BEST AVAILABLE COPY

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/02, 15/53, 15/29 | A1

(11) International Publication Number: WO 98/20113

(43) International Publication Date: 14 May 1998 (14.05.98)

(21) International Application Number: PCT/US97/20391

(22) International Filing Date: 7 November 1997 (07.11.97)

(30) Priority Data:
60/030,522 8 November 1996 (08.11.96) US
60/054,380 31 July 1997 (31.07.97) US

(71) Applicant (*for all designated States except US*): WASHINGTON STATE UNIVERSITY RESEARCH FOUNDATION [US/US]; N.E. 1615 Eastgate Boulevard, Pullman, WA 99164-1802 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): LEWIS, Norman, G. [CA/US]; 1710 N.E. Upper Drive, Pullman, WA 99163 (US). DAVIN, Laurence, B. [FR/US]; 1710 N.E. Upper Drive, Pullman, WA 99163 (US). DINKOVA-KOSTOVA, Albena, T. [BG/US]; Broadview Apartments #1025, 116 W. University Parkway, Baltimore, MD 21210 (US). FUJITA, Masayuki [JP/US]; 215 N.W. Anthony #3, Pullman, WA 99163 (US). GANG, David, R. [US/US]; 921 S.E. Klemgard Avenue, B-4, Pullman, WA 99163 (US). SARKANEN, Simo [FI/US]; 4051 – 39th Avenue S., Minneapolis, MN 55406 (US).

(74) Agent: SHELTON, Dennis, K.; Christensen O'Connor Johnson & Kindness PLLC, Suite 2800, 1420 Fifth Avenue, Seattle, WA 98101 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: RECOMBINANT PINORESINOL/LARICIRESINOL REDUCTASE, RECOMBINANT DIRIGENT PROTEIN, AND METHODS OF USE

(57) Abstract

Dirigent proteins and pinoresinol/lariciresinol reductases have been isolated from *Forsythia intermedia*, *Thuja plicata* and *Tsuga heterophylla*, together with cDNAs encoding dirigent proteins and pinoresinol/lariciresinol reductases from these species. Accordingly, isolated DNA sequences are provided which code for the expression of dirigent proteins and pinoresinol/lariciresinol reductases. In other aspects, replicable recombinant cloning vehicles are provided which code for dirigent proteins or pinoresinol/lariciresinol reductases or for a base sequence sufficiently complementary to at least a portion of dirigent protein or pinoresinol/lariciresinol reductase DNA or RNA to enable hybridization therewith (e.g., antisense dirigent protein or pinoresinol/lariciresinol reductase RNA or fragments of complementary dirigent protein or pinoresinol/lariciresinol reductase DNA which are useful as polymerase chain reaction primers or as probes for genes encoding dirigent proteins or pinoresinol/lariciresinol reductases or related genes). In yet other aspects, modified host cells are provided that have been transformed, transfected, infected and/or injected with a recombinant cloning vehicle and/or DNA sequence encoding dirigent protein or pinoresinol/lariciresinol reductase. Thus, systems and methods are provided for the recombinant expression of dirigent proteins and/or pinoresinol/lariciresinol reductases that may be used to facilitate the production, isolation and purification of significant quantities of recombinant dirigent proteins and/or pinoresinol/lariciresinol reductases for subsequent use, to obtain expression or enhanced expression of dirigent proteins and/or pinoresinol/lariciresinol reductases in plants in order to enhance, or otherwise alter, lignan biosynthesis, or may be otherwise employed for the regulation or expression of dirigent proteins and pinoresinol/lariciresinol reductases.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

RECOMBINANT PINORESINOL/LARICIRESINOL REDUCTASE, RECOMBINANT DIRIGENT PROTEIN, AND METHODS OF USE

Field of the Invention

The present invention relates to isolated dirigent proteins and pinoresinol/lariciresinol reductases from *Forsythia intermedia*, *Tsuga heterophylla* and *Thuja plicata*, to nucleic acid sequences which code for dirigent proteins and pinoresinol/lariciresinol reductases from *Forsythia intermedia*, *Tsuga heterophylla* and *Thuja plicata*, and to vectors containing the sequences, host cells containing the sequences and methods of producing recombinant pinoresinol/lariciresinol reductases, recombinant dirigent protein and their mutants.

Background of the Invention

Lignans are a large, structurally diverse, class of vascular plant metabolites having a wide range of physiological functions and pharmacologically important properties (Ayres, D.C., and Loike, J.D. in *Chemistry and Pharmacology of Natural Products*. Lignans. Chemical, Biological and Clinical Properties, Cambridge University Press, Cambridge, England (1990); Lewis et al., in Chemistry of the Amazon, Biodiversity Natural Products, and Environmental Issues, 588, (P.R. Seidl, O.R. Gottlieb and M.A.C. Kaplan) 135-167, ACS Symposium Series, Washington D.C. (1995)). Because of their pronounced antibiotic properties (Markkanen, T. et al., *Drugs Exptl. Clin. Res.* 7:711-718 (1981)), antioxidant properties (Fauré, M. et al., *Phytochemistry* 29:3773-3775 (1990); Osawa, T. et al., *Agric. Biol. Chem.* 49:3351-3352 (1985)) and antifeedant properties (Harmatha, J., and Nawrot, J., *Biochem. Syst. Ecol.* 12:95-98 (1984)), a major role of lignans in vascular plants is to help confer resistance against various opportunistic biological

pathogens and predators. Lignans have also been proposed as cytokinins (Binns, A.N. et al., *Proc. Natl. Acad. Sci. USA* 84:980-984 (1987)) and as intermediates in lignification (Rahman, M.M.A. et al., *Phytochemistry* 29:1861-1866 (1990)), suggesting a critical role in plant growth and development. It is widely held that elaboration of biochemical pathways to lignins/lignans and related substances from phenylalanine (tyrosine) was essential for the successful transition of aquatic plants to their vascular dry-land counterparts (Lewis, N.G., and Davin, L.B., in *Isoprenoids and Other Natural Products. Evolution and Function*, 562 (W.D. Nes, ed) 202-246, ACS Symposium Series: Washington, DC (1994)), some four hundred and eighty million years ago (Graham, L.E., *Origin of Land Plants*, John Wiley & Sons, Inc., New York, NY (1993)).

Based on existing chemotaxonomic data, lignans are present in "primitive" plants, such as the fern *Blechnum orientale* (Wada, H. et al., *Chem. Pharm. Bull.* 40:2099-2101 (1992)) and the hornworts, e.g., *Dendroceros japonicus* and *Megaceros flagellaris* (Takeda, R. et al., in *Bryophytes. Their Chemistry and Chemical Taxonomy*, Vol. 29 (Zinsmeister, H.D. and Mues, R. eds) pp. 201-207, Oxford University Press: New York, NY (1990); Takeda, R. et al., *Tetrahedron Lett.* 31:4159-4162 (1990)), with the latter recently being classified as originating in the Silurian period (Graham, L.E., *J. Plant Res.* 109: 241-252 (1996)). Interestingly, evolution of both gymnosperms and angiosperms was accompanied by major changes in the structural complexity and oxidative modifications of the lignans (Lewis, N.G., and Davin, L.B., in *Isoprenoids and Other Natural Products. Evolution and Function*, 562 (W. D. Nes, ed) 202-246, ACS Symposium Series: Washington, DC (1994); Gottlieb, O.R., and Yoshida, M., in *Natural Products of Woody Plants. Chemicals Extraneous to the Lignocellulosic Cell Wall* (Rowe, J.W. and Kirk, C.H. eds) pp. 439-511, Springer Verlag: Berlin (1989)). Indeed, in some species, such as Western Red Cedar (*Thuja plicata*), lignans can contribute extensively to heartwood formation/generation by enhancing the resulting heartwood color, quality, fragrance and durability.

In addition to their functions in plants, lignans also have important pharmacological roles. For example, podophyllotoxin, as its etoposide and teniposide derivatives, is an example of a plant compound that has been successfully employed as an anticancer agent (Ayres, D.C., and Loike, J.D. in *Chemistry and Pharmacology of Natural Products*. Lignans. Chemical, Biological and Clinical Properties, Cambridge University Press, Cambridge, England (1990)). Antiviral

properties have also been reported for selected lignans. For example, (-)-arctigenin (Schröder, H.C. et al., Z. *Naturforsch.* **45c**, 1215-1221 (1990)), (-)-trachelogenin (Schröder, H.C. et al., Z. *Naturforsch.* **45c**, 1215-1221 (1990)) and nordihydroguaiaretic acid (Gnabre, J.N. et al., *Proc. Natl. Acad. Sci. USA* 92:11239-11243 (1995)) are each effective against HIV due to their pronounced reverse transcriptase inhibitory activities. Some lignans, e.g., matairesinol (Nikaido, T. et al., *Chem. Pharm. Bull.* 29:3586-3592 (1981)), inhibit cAMP-phosphodiesterase, whereas others enhance cardiovascular activity, e.g., syringaresinol β-D-glucoside (Nishibe, S. et al., *Chem. Pharm. Bull.* 38:1763-1765 (1990)). There is also a high correlation between the presence, in the diet, of the "mammalian" lignans or "phytoestrogens", enterolactone and enterodiol, formed following digestion of high fiber diets, and reduced incidence rates of breast and prostate cancers (so-called chemoprevention) (Axelson, M., and Setchell, K.D.R., *FEBS Lett.* 123:337-342 (1981); Adlercreutz et al., *J. Steroid Biochem. Molec. Biol.* 41:3-8 (1992); Adlercreutz et al., *J. Steroid Biochem. Molec. Biol.* 52:97-103 (1995)). The "mammalian lignans," in turn, are considered to be derived from lignans such as matairesinol and secoisolariciresinol (Boriello et al., *J. Applied Bacteriol.*, 58:37-43 (1985)).

The biosynthetic pathways to the lignans are only now being defined, although there are no prior art reports of the isolation of enzymes or genes involved in the lignan biosynthetic pathway. Based on radiolabeling experiments with crude enzyme extracts from *Forsythia intermedia*, it was first established that entry into the 8,8'-linked lignans, which represent the most prevalent dilignol linkage known (Davin, L.B., and Lewis, N.G., in *Rec. Adv. Phytochemistry*, Vol. 26 (Stafford, H.A., and Ibrahim, R.K., eds), pp. 325-375, Plenum Press, New York, NY (1992)), occurs via stereoselective coupling of two achiral coniferyl alcohol molecules, in the form of oxygenated free radicals, to afford the furofuran lignan (+)-pinoresinol (Davin, L.B., Bedgar, D.L., Katayama, T., and Lewis, N.G., *Phytochemistry* 31:3869-3874 (1992); Paré, P.W. et al., *Tetrahedron Lett.* 35:4731-4734 (1994)) (FIGURE 1).

Bimolecular phenoxy radical coupling reactions, such as the stereoselective coupling of two achiral coniferyl alcohol molecules to afford the furofuran lignan (+)-pinoresinol, are involved in numerous biological processes. These are presumed to include lignin formation in vascular plants (M. Nose et al., *Phytochemistry* 39:71 (1995)), lignan formation in vascular plants (N.G. Lewis and L.B. Davin, *ACS Symp. Ser.* 562:202 (1994); P. W. Paré et al., *Tetrahedron Lett.* 35:4731 (1994)), suberin

formation in vascular plants (M.A. Bernards et al., *J. Biol. Chem.* 270:7382 (1995)), fruiting body development in fungi (J.D. Bu'Lock et al., *J. Chem. Soc.* 2085 (1962)), insect cuticle melanization and sclerotization (M. Miessner et al., *Helv. Chim. Acta* 74:1205 (1991); V.J. Marmaras et al., *Arch. Insect Biochem. Physiol.* 31:119 (1996)), the formation of aphid pigments (D.W. Cameron and Lord Todd, in *Organic Substances of Natural Origin. Oxidative Coupling of Phenols*, W.I. Taylor and A.R. Battersby, Eds. (Dekker, New York, 1967), Vol. 1, p.203), and the formation of algal cell wall polymers (M.A. Ragan, *Phytochemistry* 23:2029 (1984)).

In contrast to the marked regiochemical and/or stereochemical specificities observed in the biosynthesis of the foregoing lignin and lignan substances *in vivo*, all previously described chemical (J. Iqbal et al., *Chem. Rev.* 94:519 (1994)) and enzymatic (K. Freudenberg, *Science* 148:595 (1965)) bimolecular phenoxy radical coupling reactions *in vitro* have lacked strict regio- and stereospecific control. That is, if chiral centers are introduced during coupling *in vitro*, the products are racemic, and different regiochemistries can result if more than one potential coupling site is present. Thus, the ability to generate a particular enantiomeric form or a specific coupling product *in vitro* is not under explicit control. Consequently, it is inferred that a mechanism exists *in vivo* to control the regiochemistry and stereochemistry of bimolecular phenoxy radical coupling reactions leading to the formation of, for example, lignans.

In *Forsythia intermedia*, and presumably other species, (+)-pinoresinol, the product of the stereospecific coupling of two E-coniferyl alcohol molecules, undergoes sequential reduction to generate (+)-lariciresinol and then (-)-secoisolariciresinol (Katayama, T. et al., *Phytochemistry* 32:581-591 (1993); Chu, A. et al., *J. Biol. Chem.* 268:27026-27033 (1993)) (FIGURE 1). While it has hitherto been unclear whether more than one reductase is required to catalyze the sequential steps, the reductions proceed via abstraction of the pro-R hydride of NADPH, resulting in an "inversion" of configuration at both the C-7 and C-7' positions of the products, (+)-lariciresinol and (-)-secoisolariciresinol (Chu, A., et al., *J. Biol. Chem.* 268:27026-27033 (1993)). (-)-Matairesinol is subsequently formed via dehydrogenation of (-)-secoisolariciresinol, further metabolism of which presumably affords lignans such as the antiviral (-)-trachelogenin in *Ipomoea cairica* and (-)-podophyllotoxin in *Podophyllum peltatum*.

Thus, the stereospecific formation of (+)-pinoresinol and the subsequent reductive steps giving (+)-lariciresinol and (-)-secoisolariciresinol are pivotal points

in lignan metabolism, since they represent entry into the furano, dibenzylbutane, dibenzylbutyrolactone and aryltetrahydronaphthalene lignan subclasses. Additionally, it should be noted that while lignans are normally optically active, the particular enantiomer present may differ between plant species. For example, (-)-pinoresinol occurs in *Xanthoxylum ailanthoides* (Ishii et al., *Yakugaku Zasshi*, 103:279-292 (1983)), and (-)-lariciresinol is present in *Daphne tangutica* (Lin-Gen, et al., *Planta Medica*, 45:172-176 (1982)). The optical activity of a particular lignan may have important ramifications regarding biological activity. For example, (-)-trachelogenin inhibits the *in vitro* replication of HIV-1, whereas its (+)-enantiomer is much less effective (Schroder et al., *Naturforsch.* 45c:1215-1221(1990)).

Summary of the Invention

In accordance with the foregoing, in one aspect of the invention it has now been discovered that a 78-kD dirigent protein is involved in conferring stereospecificity in 8,8'-linked lignan formation. This protein has no detectable catalytically active oxidative center and apparently serves only to bind and orient coniferyl alcohol-derived free radicals, which then undergo stereoselective coupling to form (+)-pinoresinol. The formation of free-radicals, in the first instance, requires the oxidative capacity of either a nonspecific oxidase or even a non-enzymatic electron oxidant. In another aspect of the invention, it has been discovered that a single enzyme, designated pinoresinol/lariciresinol reductase, catalyzes the conversion of pinoresinol to lariciresinol and then to secoisolariciresinol. Thus, one aspect of the invention relates to isolated dirigent proteins and to isolated pinoresinol/lariciresinol reductases, such as, for example, those from *Forsythia intermedia*, *Thuja plicata* and *Tsuga heterophylla*.

In other aspects of the invention, cDNAs encoding dirigent protein from *Forsythia intermedia* (SEQ ID Nos:12 and 14), *Thuja plicata* (SEQ ID Nos:20,22,24,26,28,30,32 and 34) and *Tsuga heterophila* (SEQ ID Nos:16 and 18) have been isolated and sequenced, and the corresponding amino acid sequences have been deduced. Also, cDNAs encoding pinoresinol/lariciresinol reductase from *Forsythia intermedia* (SEQ ID Nos:47,49,51,53,55 and 57), *Thuja plicata* (SEQ ID Nos:61,63,65 and 67) and *Tsuga heterophila* (SEQ ID Nos:69 and 71) have been isolated and sequenced, and the corresponding amino acid sequences have been deduced.

Thus, the present invention relates to isolated proteins and to isolated DNA sequences which code for the expression of dirigent protein or pinoresinol/-lariciresinol reductase. In other aspects, the present invention is directed to replicable recombinant cloning vehicles comprising a nucleic acid sequence which codes for a pinoresinol/lariciresinol reductase or for a dirigent protein. The present invention is also directed to a base sequence sufficiently complementary to at least a portion of a pinoresinol/lariciresinol reductase DNA or RNA, or to at least a portion of a dirigent protein DNA or RNA, to enable hybridization therewith. The aforesaid complementary base sequences include, but are not limited to: antisense pinoresinol/lariciresinol reductase RNA; antisense dirigent protein RNA; fragments of DNA that are complementary to a pinoresinol/lariciresinol reductase DNA, or to a dirigent protein DNA, and which are therefore useful as polymerase chain reaction primers, or as probes for pinoresinol/lariciresinol reductase genes, dirigent protein genes, or related genes.

In yet other aspects of the invention, modified host cells are provided that have been transformed, transfected, infected and/or injected with a recombinant cloning vehicle and/or DNA sequence of the invention. Thus, the present invention provides for the recombinant expression of pinoresinol/lariciresinol reductases and dirigent proteins in plants, animals, microbes and in cell cultures. The inventive concepts described herein may be used to facilitate the production, isolation and purification of significant quantities of recombinant pinoresinol/lariciresinol reductase or dirigent protein, or of their enzyme products, in plants, animals, microbes or cell cultures.

Brief Description of the Drawings

The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same becomes better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1 shows the stereospecific conversion of *E*-coniferyl alcohol to (+)-pinoresinol in *Forsythia intermedia*. The stereoselectivity of this reaction is controlled by dirigent protein. (+)-Pinoresinol is then sequentially converted to (+)-lariciresinol and (-)-secoisolariciresinol by (+)-pinoresinol/(+)-lariciresinol reductase. (+)-pinoresinol, (+)-lariciresinol and (-)-secoisolariciresinol are the precursors of the furofuran, furano and dibenzylbutane families of lignans, respectively.

Detailed Description of the Preferred Embodiment

As used herein, the terms "amino acid" and "amino acids" refer to all naturally occurring L-α-amino acids or their residues. The amino acids are identified by either the single-letter or three-letter designations:

Asp	D	aspartic acid	Ile	I	isoleucine
Thr	T	threonine	Leu	L	leucine
Ser	S	serine	Tyr	Y	tyrosine
Glu	E	glutamic acid	Phe	F	phenylalanine
Pro	P	proline	His	H	histidine
Gly	G	glycine	Lys	K	lysine
Ala	A	alanine	Arg	R	arginine
Cys	C	cysteine	Trp	W	tryptophan
Val	V	valine	Gln	Q	glutamine
Met	M	methionine	Asn	N	asparagine

As used herein, the term "nucleotide" means a monomeric unit of DNA or RNA containing a sugar moiety (pentose), a phosphate and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of pentose) and that combination of base and sugar is called a nucleoside. The base characterizes the nucleotide with the four bases of DNA being adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). Inosine ("I") is a synthetic base that can be used to substitute for any of the four, naturally-occurring bases (A, C, G or T). The four RNA bases are A,G,C and uracil ("U"). The nucleotide sequences described herein comprise a linear array of nucleotides connected by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

The term "percent identity" (%I) means the percentage of amino acids or nucleotides that occupy the same relative position when two amino acid sequences, or two nucleic acid sequences, are aligned side by side.

The term "percent similarity" (%S) is a statistical measure of the degree of relatedness of two compared protein sequences. The percent similarity is calculated by a computer program that assigns a numerical value to each compared pair of amino acids based on chemical similarity (e.g., whether the compared amino acids are acidic, basic, hydrophobic, aromatic, etc.) and/or evolutionary distance as measured by the minimum number of base pair changes that would be required to convert a codon encoding one member of a pair of compared amino acids to a codon

encoding the other member of the pair. Calculations are made after a best fit alignment of the two sequences has been made empirically by iterative comparison of all possible alignments. (Henikoff, S. and Henikoff, J.G., *Proc. Nat'l Acad Sci USA* 89:10915-10919 (1992)).

"Oligonucleotide" refers to short length single or double stranded sequences of deoxyribonucleotides linked via phosphodiester bonds. The oligonucleotides are chemically synthesized by known methods and purified, for example, on polyacrylamide gels.

The term "pinoresinol/lariciresinol reductase" is used herein to mean an enzyme capable of catalyzing two reduction reactions: the reduction of pinoresinol to lariciresinol, and the reduction of lariciresinol to secoisolariciresinol. The products of these reactions, lariciresinol and secoisolariciresinol, can be either the (+)- or (-)-enantiomers.

The term "dirigent protein" is used herein to mean a protein capable of guiding a bimolecular phenoxy radical coupling reaction thereby determining the stereochemistry and regiochemistry of the product of the reaction and/or its polymeric derivatives.

The terms "alteration", "amino acid sequence alteration", "variant" and "amino acid sequence variant" refer to dirigent protein or pinoresinol/lariciresinol reductase molecules with some differences in their amino acid sequences as compared to the corresponding native dirigent protein or pinoresinol/lariciresinol reductase. Ordinarily, the variants will possess at least about 70% homology with the corresponding, native dirigent protein or pinoresinol/lariciresinol reductase, and preferably they will be at least about 80% homologous with the corresponding, native dirigent protein or pinoresinol/lariciresinol reductase. The amino acid sequence variants of dirigent protein or pinoresinol/lariciresinol reductase falling within this invention possess substitutions, deletions, and/or insertions at certain positions. Sequence variants of dirigent protein or pinoresinol/lariciresinol reductase may be used to attain desired enhanced or reduced enzymatic activity, modified regiochemistry or stereochemistry, or altered substrate utilization or product distribution.

Substitutional dirigent protein variants or pinoresinol/lariciresinol reductase variants are those that have at least one amino acid residue in the corresponding native dirigent protein sequence or pinoresinol/lariciresinol reductase sequence removed and a different amino acid inserted in its place at the same position. The

substitutions may be single, where only one amino acid in the molecule has been substituted, or they may be multiple, where two or more amino acids have been substituted in the same molecule. Substantial changes in the activity of the dirigent protein or pinoresinol/lariciresinol reductase molecule may be obtained by substituting an amino acid with a side chain that is significantly different in charge and/or structure from that of the native amino acid. This type of substitution would be expected to affect the structure of the polypeptide backbone and/or the charge or hydrophobicity of the molecule in the area of the substitution.

Moderate changes in the activity of the dirigent protein or pinoresinol/-lariciresinol reductase molecule would be expected by substituting an amino acid with a side chain that is similar in charge and/or structure to that of the native molecule. This type of substitution, referred to as a conservative substitution, would not be expected to substantially alter either the structure of the polypeptide backbone or the charge or hydrophobicity of the molecule in the area of the substitution.

Insertional dirigent protein variants or pinoresinol/lariciresinol reductase variants are those with one or more amino acids inserted immediately adjacent to an amino acid at a particular position in the native dirigent protein or pinoresinol/-lariciresinol reductase molecule. Immediately adjacent to an amino acid means connected to either the α-carboxy or α-amino functional group of the amino acid. The insertion may be one or more amino acids. Ordinarily, the insertion will consist of one or two conservative amino acids. Amino acids similar in charge and/or structure to the amino acids adjacent to the site of insertion are defined as conservative. Alternatively, this invention includes insertion of an amino acid with a charge and/or structure that is substantially different from the amino acids adjacent to the site of insertion.

Deletional variants are those where one or more amino acids in the native dirigent protein or pinoresinol/lariciresinol reductase molecule have been removed. Ordinarily, deletional variants will have one or two amino acids deleted in a particular region of the dirigent protein or pinoresinol/lariciresinol reductase molecule.

The term "antisense" or "antisense RNA" or "antisense nucleic acid" is used herein to mean a nucleic acid molecule that is complementary to all or part of a messenger RNA molecule. Antisense nucleic acid molecules are typically used to inhibit the expression, *in vivo*, of complementary, expressed messenger RNA molecules.

The terms "biological activity", "biologically active", "activity" and "active" when used with reference to a pinoresinol/lariciresinol reductase molecule refer to the ability of the pinoresinol/lariciresinol reductase molecule to reduce pinoresinol and lariciresinol to yield lariciresinol and secoisolariciresinol, respectively, as measured in an enzyme activity assay, such as the assay described in Example 8 below.

The terms "biological activity", "biologically active", "activity" and "active" when used with reference to a dirigent protein refer to the ability of the dirigent protein to guide a bimolecular phenoxy radical coupling reaction thereby determining the stereochemistry and regiochemistry of the product of the reaction and of its polymeric derivatives.

Amino acid sequence variants of dirigent protein or pinoresinol/lariciresinol reductase may have desirable altered biological activity including, for example, altered reaction kinetics, substrate utilization, product distribution or other characteristics such as regiochemistry and stereochemistry.

The terms "DNA sequence encoding", "DNA encoding" and "nucleic acid encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the translated polypeptide chain. The DNA sequence thus codes for the amino acid sequence.

The terms "replicable expression vector" and "expression vector" refer to a piece of DNA, usually double-stranded, which may have inserted into it a piece of foreign DNA. Foreign DNA is defined as heterologous DNA, which is DNA not naturally found in the host. The vector is used to transport the foreign or heterologous DNA into a suitable host cell. Once in the host cell, the vector can replicate independently of or coincidentally with the host chromosomal DNA, and several copies of the vector and its inserted (foreign) DNA may be generated. In addition, the vector contains the necessary elements that permit translating the foreign DNA into a polypeptide. Many molecules of the polypeptide encoded by the foreign DNA can thus be rapidly synthesized.

The terms "transformed host cell," "transformed" and "transformation" refer to the introduction of DNA into a cell. The cell is termed a "host cell", and it may be a prokaryotic or a eukaryotic cell. Typical prokaryotic host cells include various strains of *E. coli*. Typical eukaryotic host cells are plant cells, such as maize cells, yeast cells, insect cells or animal cells. The introduced DNA is usually in the form of

a vector containing an inserted piece of DNA. The introduced DNA sequence may be from the same species as the host cell or from a different species from the host cell, or it may be a hybrid DNA sequence, containing some foreign DNA and some DNA derived from the host species.

In accordance with the present invention, cDNAs encoding dirigent protein and pinoresinol/lariciresinol reductase from *Forsythia intermedia, Thuja plicata* and *Tsuga heterophylla* were isolated, sequenced and expressed in the following manner.

With respect to the cDNAs encoding dirigent protein from *Forsythia intermedia*, an empirically-determined purification protocol was developed to isolate the *Forsythia* dirigent protein. This procedure yielded at least six isoforms of the dirigent protein. Amino acid sequencing of the amino terminus of each of these isoforms revealed that the sequence of each isoform was identical. Sequencing of the N-terminus of a mixture of these isoforms yielded a 28 amino acid sequence (SEQ ID No:1). Tryptic digestion of a mixture of these isoforms yielded six peptide fragments which were purified in sufficient quantity to permit sequencing SEQ ID Nos:2-7.

A primer designated PSINT1 (SEQ ID No:8) was synthesized based on the sequence of amino acids 9 to 15 of the N-terminal peptide (SEQ ID No:1). A primer designated PSI1R (SEQ ID No:9) was synthesized based on the sequence of amino acids 3 to 9 of the internal peptide sequence set forth in (SEQ ID No:2). A primer designated PSI2R (SEQ ID No:10) was synthesized based on the sequence of amino acids 13 to 20 of the internal peptide sequence set forth in (SEQ ID No:2). A primer designated PSI7R (SEQ ID No:11) was synthesized based on the sequence of amino acids 6 to 12 of the internal peptide sequence set forth in (SEQ ID No:3).

Forsythia total RNA was isolated by means of a protocol adapted from a method specifically designed for woody tissues which contain a large concentration of polyphenols. Poly A+ RNA was isolated and a cDNA library constructed using standard means. A PCR reaction utilizing primers PSINT1 (SEQ ID No:8) and one of PSI7R, (SEQ ID No:11) PSI2R (SEQ ID No:10) or PSI1R (SEQ ID No:9), together with an aliquot of *Forsythia* cDNA as substrate, each yielded a single cDNA band of ~370 bp, ~155 bp and ~125 bp, respectively. The ~370 bp product of the PSINT1 (SEQ ID NO:8)-PSI7R (SEQ ID No:11) reaction was amplified by PCR and utilized as a probe to screen approximatley 600,000 PFU of a *Forsythia intermedia* cDNA library. Two distinct cDNAs were identified, called pPSDFi1 (SEQ ID No:12) and pPSDFi2 (SEQ ID No:14). The cDNA insert encoding dirigent

protein was excised from plasmid pPSDFi1 and cloned into the baculovirus transfer vector pBlueBac4. The resulting construct was used to transform *Spodoptera frugiperda* from which functional dirigent protein was purified.

With respect to the cloning of dirigent protein from *Thuja plicata* and *Tsuga heterophylla*, the *Forsythia* cDNAs were used as probes to isolate two dirigent protein clones from *Tsuga heterophylla* (SEQ ID Nos:16, 18), and eight dirigent protein cDNA clones from *Thuja plicata* (SEQ ID Nos:20, 22, 24, 26, 28, 30, 32, 34).

With respect to the cDNAs encoding (+)-pinoresinol/(+)-lariciresinol reductase from *Forsythia intermedia*, an empirically-determined purification protocol, consisting of eight chromatographic steps, was developed to isolate the *Forsythia* (+)-pinoresinol/(+)-lariciresinol reductase protein. This procedure yielded two isoforms of (+)-pinoresinol/(+)-lariciresinol reductase which were both capable of catalyzing the reduction of (+)-pinoresinol and (+)-lariciresinol. Sequencing of the N-terminus of each of these isoforms yielded an identical 30 amino acid sequence (SEQ ID No:36). Tryptic digestion of a mixture of both of these isoforms yielded four peptide fragments which were purified in sufficient quantity to permit sequencing (SEQ ID Nos:37-40). Additionally, cyanogen bromide cleavage of a mixture of both of these isoforms yielded three peptide fragments which were purified in sufficient quantity to permit sequencing (SEQ ID Nos:41-43).

A primer designated PLRN5 (SEQ ID No:44) was synthesized based on the sequence of amino acids 7 to 13 of the N-terminal peptide (SEQ ID No:36). A primer designated PLR14R (SEQ ID No:45) was synthesized based on the sequence of amino acids 2 to 8 of the internal peptide sequence set forth in SEQ ID No:37. A primer designated PLR15R (SEQ ID No:46) was synthesized based on the sequence of amino acids 9 to 15 of the internal peptide sequence set forth in SEQ ID No:37. The sequence of amino acids 9 to 15 of the internal peptide sequence set forth in SEQ ID No:37, upon which the sequence of primer PLR15R (SEQ ID No:46) was based, also corresponded to the sequence of amino acids 4 to 10 of the cyanogen bromide-generated, internal fragment set forth in SEQ ID No:41.

Forsythia total RNA was isolated by means of a protocol adapted from a method specifically designed for woody tissues which contain a large concentration of polyphenols. Poly A+ RNA was isolated and a cDNA library constructed using standard means. A PCR reaction utilizing primers PLRN5 (SEQ ID No:44) and either PLR14R (SEQ ID No:45) or PLR15R (SEQ ID No:46), together with an

aliquot of *Forsythia* cDNA as substrate, yielded two, amplified bands of 380 bp and 400 bp. One 400 bp cDNA insert was utilized as a probe with which to screen the *Forsythia* cDNA library. The 400 bp probe corresponded to bases 22 to 423 of SEQ ID No:47. Six cDNA clones were isolated and sequenced (SEQ ID Nos:47, 49, 51, 53, 55, 57). The clones shared a common coding region, many had a different 5'-untranslated region and the 3'-untranslated region of each terminated at a different point. One of these cDNAs (SEQ ID No:47), expressed as a β-galactosidase fusion protein in *E. coli*, catalyzed the same enantiomer-specific reactions as the native plant protein.

With respect to the cloning of (+)-pinoresinol/(+)-lariciresinol reductase and (-)-pinoresinol/(-)-lariciresinol reductase from *Thuja plicata*, cDNA was synthesized and utilized as a template in a PCR reaction in which the primers were a 3' linker-primer (SEQ ID No:59) and a 5' primer, designated CR6-NT, (SEQ ID No:60). At least two bands of the expected length (1.2 kb) were generated and cloned into a plasmid vector. One clone, designated plr-Tp1, (SEQ ID No:61) was completely sequenced and expressed as a β-galactosidase fusion protein in *E. coli*. plr-Tp1 encodes a (-)-pinoresinol/(-)-lariciresinol reductase.

The cDNA insert of clone plr-Tp1 was used to screen the *T. plicata* cDNA library and identified an additional, unique clone, designated plr-Tp2, (SEQ ID No:63). plr-Tp2 has high homology to plr-Tp1 but encodes a (+)-pinoresinol/(+)-lariciresinol reductase. The cDNA insert of clone plr-Tp1 was used to screen the *T. plicata* cDNA library and identify an additional two pinoresinol/lariciresinol reductase cDNAs (SEQ ID Nos:65, 67).

Two cDNAs encoding pinoresinol/lariciresinol reductases from *Tsuga heterophylla* (SEQ ID Nos:69, 71) were isolated by screening a *Tsuga heterophylla* cDNA library with the plr-Tp1 cDNA insert.

The isolation of cDNAs encoding dirigent proteins, (+)-pinoresinol/-(+)-lariciresinol reductase and (-)-pinoresinol/(-)-lariciresinol reductase permits the development of an efficient expression system for these functional enzymes; provides useful tools for examining the developmental regulation of lignan biosynthesis and permits the isolation of other dirigent proteins and pinoresinol/lariciresinol reductases. The isolation of the dirigent protein and pinoresinol/lariciresinol reductase cDNAs also permits the transformation of a wide range of organisms in order to enhance or modify lignan biosynthesis.

The proteins and nucleic acids of the present invention can be utilized to predetermine the stereochemistry, regiochemistry, or both, of the products of bimolecular phenoxy coupling reactions, such as the furofuran, furano and dibenzylbutane lignans. By way of non-limiting examples, the proteins and nucleic acids of the present invention can be utilized to: elevate or otherwise alter the levels of health-protecting lignans, such as podophyllotoxin, in plant species, including but not limited to vegetables, grains and fruits, and to food items incorporating material derived from such genetically altered plants; genetically alter plant species to provide an abundant, natural supply of lignans useful for a variety of purposes, for example as neutriceuticals and dietary supplements; to genetically alter living organisms to produce an abundant supply of optically pure lignans having desirable biological properties, for example (-)-arctigenin which possesses antiviral properties. In particular, characterization of the dirigent protein binding site and mechanism of action permits the development of synthetic proteins consisting of an array of dirigent protein binding sites which serve as templates for stereochemically-controlled polymeric assembly.

N-terminal transport sequences well known in the art (see, e.g., von Heijne, G. et al., *Eur. J. Biochem* 180:535-545 (1989); Stryer, *Biochemistry* W.H. Freeman and Company, New York, NY, p. 769 (1988)) may be employed to direct the dirigent protein or pinoresinol/lariciresinol reductase to a variety of cellular or extracellular locations.

Sequence variants of wild-type dirigent protein clones and pinoresinol/-lariciresinol clones that can be produced by deletions, substitutions, mutations and/or insertions are intended to be within the scope of the invention except insofar as limited by the prior art. Dirigent protein or pinoresinol/lariciresinol reductase amino acid sequence variants may be constructed by mutating the DNA sequence that encodes wild-type dirigent protein or wild-type pinoresinol/lariciresinol reductase, such as by using techniques commonly referred to as site-directed mutagenesis. Various polymerase chain reaction (PCR) methods now well known in the field, such as a two primer system like the Transformer Site-Directed Mutagenesis kit from Clontech, may be employed for this purpose.

Following denaturation of the target plasmid in this system, two primers are simultaneously annealed to the plasmid; one of these primers contains the desired site-directed mutation, the other contains a mutation at another point in the plasmid resulting in elimination of a restriction site. Second strand synthesis is then carried

out, tightly linking these two mutations, and the resulting plasmids are transformed into a *mutS* strain of *E. coli*. Plasmid DNA is isolated from the transformed bacteria, restricted with the relevant restriction enzyme (thereby linearizing the unmutated plasmids), and then retransformed into *E. coli*. This system allows for generation of mutations directly in an expression plasmid, without the necessity of subcloning or generation of single-stranded phagemids. The tight linkage of the two mutations and the subsequent linearization of unmutated plasmids results in high mutation efficiency and allows minimal screening. Following synthesis of the initial restriction site primer, this method requires the use of only one new primer type per mutation site. Rather than prepare each positional mutant separately, a set of "designed degenerate" oligonucleotide primers can be synthesized in order to introduce all of the desired mutations at a given site simultaneously. Transformants can be screened by sequencing the plasmid DNA through the mutagenized region to identify and sort mutant clones. Each mutant DNA can then be restricted and analyzed by electrophoresis on Mutation Detection Enhancement gel (J.T. Baker) to confirm that no other alterations in the sequence have occurred (by band shift comparison to the unmutagenized control).

The verified mutant duplexes can be cloned into a replicable expression vector, if not already cloned into a vector of this type, and the resulting expression construct used to transform *E. coli*, such as strain *E. coli* BL21(DE3)pLysS, for high level production of the mutant protein, and subsequent purification thereof. The method of FAB-MS mapping can be employed to rapidly check the fidelity of mutant expression. This technique provides for sequencing segments throughout the whole protein and provides the necessary confidence in the sequence assignment. In a mapping experiment of this type, protein is digested with a protease (the choice will depend on the specific region to be modified since this segment is of prime interest and the remaining map should be identical to the map of unmutagenized protein). The set of cleavage fragments is fractionated by microbore HPLC (reversed phase or ion exchange, again depending on the specific region to be modified) to provide several peptides in each fraction, and the molecular weights of the peptides are determined by FAB-MS. The masses are then compared to the molecular weights of peptides expected from the digestion of the predicted sequence, and the correctness of the sequence quickly ascertained. Since this mutagenesis approach to protein modification is directed, sequencing of the altered peptide should not be necessary if the MS agrees with prediction. If necessary to verify a changed residue,

CAD-tandem MS/MS can be employed to sequence the peptides of the mixture in question, or the target peptide purified for subtractive Edman degradation or carboxypeptidase Y digestion depending on the location of the modification.

In the design of a particular site directed mutant, it is generally desirable to first make a non-conservative substitution (e.g., Ala for Cys, His or Glu) and determine if activity is greatly impaired as a consequence. The properties of the mutagenized protein are then examined with particular attention to the kinetic parameters of K_m and k_{cat} as sensitive indicators of altered function, from which changes in binding and/or catalysis *per se* may be deduced by comparison to the native enzyme. If the residue is by this means demonstrated to be important by activity impairment, or knockout, then conservative substitutions can be made, such as Asp for Glu to alter side chain length, Ser for Cys, or Arg for His. For hydrophobic segments, it is largely size that will be altered, although aromatics can also be substituted for alkyl side chains. Changes in the normal product distribution can indicate which step(s) of the reaction sequence have been altered by the mutation.

Other site directed mutagenesis techniques may also be employed with the nucleotide sequences of the invention. For example, restriction endonuclease digestion of DNA followed by ligation may be used to generate dirigent protein or pinoresinol/lariciresinol reductase deletion variants, as described in Section 15.3 of Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, New York, NY (1989)). A similar strategy may be used to construct insertion variants, as described in Section 15.3 of Sambrook et al., *supra*.

Oligonucleotide-directed mutagenesis may also be employed for preparing substitution variants of this invention. It may also be used to conveniently prepare the deletion and insertion variants of this invention. This technique is well known in the art as described by Adelman et al. (*DNA* 2:183 (1983)). Generally, oligonucleotides of at least 25 nucleotides in length are used to insert, delete or substitute two or more nucleotides in the dirigent protein gene or pinoresinol/-lariciresinol reductase gene. An optimal oligonucleotide will have 12 to 15 perfectly matched nucleotides on either side of the nucleotides coding for the mutation. To mutagenize the wild-type dirigent protein or wild-type pinoresinol/lariciresinol reductase, the oligonucleotide is annealed to the single-stranded DNA template molecule under suitable hybridization conditions. A DNA polymerizing enzyme, usually the Klenow fragment of *E. coli* DNA polymerase I, is then added. This

enzyme uses the oligonucleotide as a primer to complete the synthesis of the mutation-bearing strand of DNA. Thus, a heteroduplex molecule is formed such that one strand of DNA encodes the wild-type dirigent protein or pinoresinol/lariciresinol reductase inserted in the vector, and the second strand of DNA encodes the mutated form of dirigent protein or pinoresinol/lariciresinol reductase inserted into the same vector. This heteroduplex molecule is then transformed into a suitable host cell.

Mutants with more than one amino acid substituted may be generated in one of several ways. If the amino acids are located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If however, the amino acids are located some distance from each other (separated by more than ten amino acids, for example) it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed. In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions.

An alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: wild-type dirigent protein or pinoresinol/lariciresinol reductase DNA is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third round of mutagenesis, and so on.

Eukaryotic expression systems may be utilized for dirigent protein or pinoresinol/lariciresinol reductase production since they are capable of carrying out any required posttranslational modifications and of directing the enzyme to the proper membrane location. A representative eukaryotic expression system for this purpose uses the recombinant baculovirus, *Autographa californica* nuclear polyhedrosis virus (AcNPV; M.D. Summers and G.E. Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures* (1986); Luckow et al.,

Bio-technology 6:47-55 (1987)) for expression of the dirigent protein or pinoresinol/lariciresinol reductases of the invention. Infection of insect cells (such as cells of the species *Spodoptera frugiperda*) with the recombinant baculoviruses allows for the production of large amounts of the dirigent protein or pinoresinol/lariciresinol reductase protein. In addition, the baculovirus system has other important advantages for the production of recombinant dirigent protein or pinoresinol/lariciresinol reductase. For example, baculoviruses do not infect humans and can therefore be safely handled in large quantities. In the baculovirus system, a DNA construct is prepared including a DNA segment encoding dirigent protein or pinoresinol/lariciresinol reductase and a vector. The vector may comprise the polyhedron gene promoter region of a baculovirus, the baculovirus flanking sequences necessary for proper cross-over during recombination (the flanking sequences comprise about 200-300 base pairs adjacent to the promoter sequence) and a bacterial origin of replication which permits the construct to replicate in bacteria. The vector is constructed so that (i) the DNA segment is placed adjacent (or operably-linked or "downstream" or "under the control of") to the polyhedron gene promoter and (ii) the promoter/pinoresinol/lariciresinol reductase, or promoter/-dirigent protein, combination is flanked on both sides by 200-300 base pairs of baculovirus DNA (the flanking sequences).

To produce a dirigent protein DNA construct, or a pinoresinol/lariciresinol reductase DNA construct, a cDNA clone encoding a full length dirigent protein or pinoresinol/lariciresinol reductase is obtained using methods such as those described herein. The DNA construct is contacted in a host cell with baculovirus DNA of an appropriate baculovirus (that is, of the same species of baculovirus as the promoter encoded in the construct) under conditions such that recombination is effected. The resulting recombinant baculoviruses encode the full dirigent protein or pinoresinol/lariciresinol reductase. For example, an insect host cell can be cotransfected or transfected separately with the DNA construct and a functional baculovirus. Resulting recombinant baculoviruses can then be isolated and used to infect cells to effect production of dirigent protein or pinoresinol/lariciresinol reductase. Host insect cells include, for example, *Spodoptera frugiperda* cells. Insect host cells infected with a recombinant baculovirus of the present invention are then cultured under conditions allowing expression of the baculovirus-encoded dirigent protein or pinoresinol/lariciresinol reductase. Recombinant protein thus produced is then extracted from the cells using methods known in the art.

Other eukaryotic microbes such as yeasts may also be used to practice this invention. The baker's yeast *Saccharomyces cerevisiae*, is a commonly used yeast, although several other strains are available. The plasmid YRp7 (Stinchcomb et al., *Nature* 282:39 (1979); Kingsman et al., *Gene* 7:141 (1979); Tschemper et al., *Gene* 10:157 (1980)) is commonly used as an expression vector in *Saccharomyces*. This plasmid contains the trp1 gene that provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, such as strains ATCC No. 44,076 and PEP4-1 (Jones, *Genetics* 85:12 (1977)). The presence of the trp1 lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan. Yeast host cells are generally transformed using the polyethylene glycol method, as described by Hinnen (*Proc. Natl. Acad. Sci. USA* 75:1929 (1978)). Additional yeast transformation protocols are set forth in Gietz et al., *N.A.R.* 20(17):1425 (1992); Reeves et al., *FEMS* 99:193-197 (1992).

Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem.* 255:2073 (1980)) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg.* 7:149 (1968); Holland et al., *Biochemistry* 17:4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose-phosphate isomerase, phosphoglucose isomerase, and glucokinase. In the construction of suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination. Other promoters that have the additional advantage of transcription controlled by growth conditions are the promoter region for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Any plasmid vector containing yeast-compatible promoter, origin of replication and termination sequences is suitable.

Cell cultures derived from multicellular organisms, such as plants, may be used as hosts to practice this invention. Transgenic plants can be obtained, for example, by transferring plasmids that encode pinoresinol/lariciresinol reductase, and/or dirigent protein, and a selectable marker gene, e.g., the kan gene encoding

resistance to kanamycin, into *Agrobacterium tumifaciens* containing a helper Ti plasmid as described in Hoeckema et al., *Nature* 303:179-181 (1983) and culturing the *Agrobacterium* cells with leaf slices of the plant to be transformed as described by An et al., *Plant Physiology* 81:301-305 (1986). Transformation of cultured plant host cells is normally accomplished through *Agrobacterium tumifaciens*, as described above. Cultures of mammalian host cells and other host cells that do not have rigid cell membrane barriers are usually transformed using the calcium phosphate method as originally described by Graham and Van der Eb (*Virology* 52:546 (1978)) and modified as described in Sections 16.32-16.37 of Sambrook et al., *supra*. However, other methods for introducing DNA into cells such as Polybrene (Kawai and Nishizawa, *Mol. Cell. Bio l.* 4:1172 (1984)), protoplast fusion (Schaffner, *Proc. Natl. Acad. Sci. USA* 77:2163 (1980)), electroporation (Neumann et al., *EMBO J.* 1:841 (1982)), and direct microinjection into nuclei (Capecchi, *Cell* 22:479 (1980)) may also be used. Additionally, animal transformation strategies are reviewed in Monastersky G.M. and Robl, J.M., *Strategies in Transgenic Animal Science*, ASM Press, Washington, D.C. (1995). Transformed plant calli may be selected through the selectable marker by growing the cells on a medium containing, e.g., kanamycin, and appropriate amounts of phytohormone such as naphthalene acetic acid and benzyladenine for callus and shoot induction. The plant cells may then be regenerated and the resulting plants transferred to soil using techniques well known to those skilled in the art.

In addition, a gene regulating pinoresinol/lariciresinol reductase production, or dirigent protein production, can be incorporated into the plant along with a necessary promoter which is inducible. In the practice of this embodiment of the invention, a promoter that only responds to a specific external or internal stimulus is fused to the target cDNA. Thus, the gene will not be transcribed except in response to the specific stimulus. As long as the gene is not being transcribed, its gene product is not produced.

An illustrative example of a responsive promoter system that can be used in the practice of this invention is the glutathione-S-transferase (GST) system in maize. GSTs are a family of enzymes that can detoxify a number of hydrophobic electrophilic compounds that often are used as pre-emergent herbicides (Weigand et al., *Plant Molecular Biology* 7:235-243 (1986)). Studies have shown that the GSTs are directly involved in causing this enhanced herbicide tolerance. This action is primarily mediated through a specific 1.1 kb mRNA transcription

product. In short, maize has a naturally occurring quiescent gene already present that can respond to external stimuli and that can be induced to produce a gene product. This gene has previously been identified and cloned. Thus, in one embodiment of this invention, the promoter is removed from the GST responsive gene and attached to a pinoresinol/lariciresinol reductase gene, or a dirigent protein gene, that previously has had its native promoter removed. This engineered gene is the combination of a promoter that responds to an external chemical stimulus and a gene responsible for successful production of pinoresinol/lariciresinol reductase or dirigent protein.

In addition to the methods described above, several methods are known in the art for transferring cloned DNA into a wide variety of plant species, including gymnosperms, angiosperms, monocots and dicots (see, e.g., Glick and Thompson, eds., *Methods in Plant Molecular Biology*, CRC Press, Boca Raton, Florida (1993)). Representative examples include electroporation-facilitated DNA uptake by protoplasts (Rhodes et al., *Science* 240(4849):204-207 (1988)); treatment of protoplasts with polyethylene glycol (Lyznik et al., *Plant Molecular Biology* 13:151-161 (1989)); and bombardment of cells with DNA laden microprojectiles (Klein et al., *Plant Physiol.* 91:440-444 (1989) and Boynton et al., *Science* 240(4858):1534-1538 (1988)). Numerous methods now exist, for example, for the transformation of cereal crops (see, e.g., McKinnon, G.E. and Henry, R.J., *J. Cereal Science*, 22(3):203-210 (1995); Mendel, R.R. and Teeri, T.H., *Plant and Microbial Biotechnology Research Series*, 3:81-98, Cambridge University Press (1995); McElroy, D. and Brettell, R.I.S., *Trends in Biotechnology*, 12(2):62-68 (1994); Christou et al., *Trends in Biotechnology*, 10(7):239-246 (1992); Christou, P. and Ford, T.L., *Annals of Botany*, 75(5): 449-454 (1995); Park et al., *Plant Molecular Biology*, 32(6):1135-1148 (1996); Altpeter et al., *Plant Cell Reports*, 16:12-17 (1996)). Additionally, plant transformation strategies and techniques are reviewed in Birch, R.G., *Ann Rev Plant Phys Plant Mol Biol* 48:297 (1997); Forester et al., *Exp. Agric.* 33:15-33 (1997). Minor variations make these technologies applicable to a broad range of plant species.

Each of these techniques has advantages and disadvantages. In each of the techniques, DNA from a plasmid is genetically engineered such that it contains not only the gene of interest, but also selectable and screenable marker genes. A selectable marker gene is used to select only those cells that have integrated copies of the plasmid (the construction is such that the gene of interest and the selectable and

screenable genes are transferred as a unit). The screenable gene provides another check for the successful culturing of only those cells carrying the genes of interest. A commonly used selectable marker gene is neomycin phosphotransferase II (NPT II). This gene conveys resistance to kanamycin, a compound that can be added directly to the growth media on which the cells grow. Plant cells are normally susceptible to kanamycin and, as a result, die. The presence of the NPT II gene overcomes the effects of the kanamycin and each cell with this gene remains viable. Another selectable marker gene which can be employed in the practice of this invention is the gene which confers resistance to the herbicide glufosinate (Basta). A screenable gene commonly used is the β-glucuronidase gene (GUS). The presence of this gene is characterized using a histochemical reaction in which a sample of putatively transformed cells is treated with a GUS assay solution. After an appropriate incubation, the cells containing the GUS gene turn blue. Preferably, the plasmid will contain both selectable and screenable marker genes.

The plasmid containing one or more of these genes is introduced into either plant protoplasts or callus cells by any of the previously mentioned techniques. If the marker gene is a selectable gene, only those cells that have incorporated the DNA package survive under selection with the appropriate phytotoxic agent. Once the appropriate cells are identified and propagated, plants are regenerated. Progeny from the transformed plants must be tested to insure that the DNA package has been successfully integrated into the plant genome.

Mammalian host cells may also be used in the practice of the invention. Examples of suitable mammalian cell lines include monkey kidney CVI line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line 293S (Graham et al., *J. Gen. Virol.* 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells (Urlab and Chasin, *Proc. Natl. Acad. Sci USA* 77:4216 (1980)); mouse sertoli cells (TM4, Mather, *Biol. Reprod.* 23:243 (1980)); monkey kidney cells (CVI-76, ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor cells (MMT 060562, ATCC CCL 51); rat hepatoma cells (HTC, MI.54, Baumann et al., *J. Cell Biol.* 85:1 (1980)); and TRI cells (Mather et al., *Annals N.Y. Acad. Sci.* 383:44 (1982)). Expression vectors for these cells ordinarily include (if necessary) DNA sequences

for an origin of replication, a promoter located in front of the gene to be expressed, a ribosome binding site, an RNA splice site, a polyadenylation site, and a transcription terminator site.

Promoters used in mammalian expression vectors are often of viral origin. These viral promoters are commonly derived from polyoma virus, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The SV40 virus contains two promoters that are termed the early and late promoters. These promoters are particularly useful because they are both easily obtained from the virus as one DNA fragment that also contains the viral origin of replication (Fiers et al., *Nature* 273:113 (1978)). Smaller or larger SV40 DNA fragments may also be used, provided they contain the approximately 250-bp sequence extending from the HindIII site toward the BglI site located in the viral origin of replication.

Alternatively, promoters that are naturally associated with the foreign gene (homologous promoters) may be used provided that they are compatible with the host cell line selected for transformation.

An origin of replication may be obtained from an exogenous source, such as SV40 or other virus (e.g., Polyoma, Adeno, VSV, BPV) and inserted into the cloning vector. Alternatively, the origin of replication may be provided by the host cell chromosomal replication mechanism. If the vector containing the foreign gene is integrated into the host cell chromosome, the latter is often sufficient.

The use of a secondary DNA coding sequence can enhance production levels of pinoresinol/lariciresinol reductase or dirigent protein in transformed cell lines. The secondary coding sequence typically comprises the enzyme dihydrofolate reductase (DHFR). The wild-type form of DHFR is normally inhibited by the chemical methotrexate (MTX). The level of DHFR expression in a cell will vary depending on the amount of MTX added to the cultured host cells. An additional feature of DHFR that makes it particularly useful as a secondary sequence is that it can be used as a selection marker to identify transformed cells. Two forms of DHFR are available for use as secondary sequences, wild-type DHFR and MTX-resistant DHFR. The type of DHFR used in a particular host cell depends on whether the host cell is DHFR deficient (such that it either produces very low levels of DHFR endogenously, or it does not produce functional DHFR at all). DHFR-deficient cell lines such as the CHO cell line described by Urlaub and Chasin, *supra*, are transformed with wild-type DHFR coding sequences. After transformation, these DHFR-deficient cell lines express functional DHFR and are capable of growing in a

culture medium lacking the nutrients hypoxanthine, glycine and thymidine. Nontransformed cells will not survive in this medium.

The MTX-resistant form of DHFR can be used as a means of selecting for transformed host cells in those host cells that endogenously produce normal amounts of functional DHFR that is MTX sensitive. The CHO-KI cell line (ATCC No. CL 61) possesses these characteristics, and is thus a useful cell line for this purpose. The addition of MTX to the cell culture medium will permit only those cells transformed with the DNA encoding the MTX-resistant DHFR to grow. The nontransformed cells will be unable to survive in this medium.

Prokaryotes may also be used as host cells for the initial cloning steps of this invention. They are particularly useful for rapid production of large amounts of DNA, for production of single-stranded DNA templates used for site-directed mutagenesis, for screening many mutants simultaneously, and for DNA sequencing of the mutants generated. Suitable prokaryotic host cells include *E. coli* K12 strain 294 (ATCC No. 31,446), *E. coli* strain W3110 (ATCC No. 27,325) *E. coli* X1776 (ATCC No. 31,537), and *E. coli* B; however many other strains of *E. coli*, such as HB101, JM101, NM522, NM538, NM539, and many other species and genera of prokaryotes including bacilli such as *Bacillus subtilis*, other enterobacteriaceae such as *Salmonella typhimurium* or *Serratia marcesans*, and various *Pseudomonas* species may all be used as hosts. Prokaryotic host cells or other host cells with rigid cell walls are preferably transformed using the calcium chloride method as described in section 1.82 of Sambrook et al., *supra*. Alternatively, electroporation may be used for transformation of these cells. Prokaryote transformation techniques are set forth in Dower, W. J., in Genetic Engineering, Principles and Methods, 12:275-296, Plenum Publishing Corp. (1990); Hanahan et al., *Meth. Enzymol.*, 204:63 (1991).

As a representative example, cDNA sequences encoding dirigent protein or pinoresinol/lariciresinol reductase may be transferred to the $(His)_6$•Tag pET vector commercially available (from Novagen) for overexpression in *E. coli* as heterologous host. This pET expression plasmid has several advantages in high level heterologous expression systems. The desired cDNA insert is ligated in frame to plasmid vector sequences encoding six histidines followed by a highly specific protease recognition site (thrombin) that are joined to the amino terminus codon of the target protein. The histidine "block" of the expressed fusion protein promotes very tight binding to immobilized metal ions and permits rapid purification of the recombinant protein by immobilized metal ion affinity chromatography. The histidine leader sequence is

then cleaved at the specific proteolysis site by treatment of the purified protein with thrombin, and the dirigent protein or pinoresinol/lariciresinol reductase eluted. This overexpression-purification system has high capacity, excellent resolving power and is fast, and the chance of a contaminating *E. coli* protein exhibiting similar binding behavior (before and after thrombin proteolysis) is extremely small.

As will be apparent to those skilled in the art, any plasmid vectors containing replicon and control sequences that are derived from species compatible with the host cell may also be used in the practice of the invention. The vector usually has a replication site, marker genes that provide phenotypic selection in transformed cells, one or more promoters, and a polylinker region containing several restriction sites for insertion of foreign DNA. Plasmids typically used for transformation of *E. coli* include pBR322, pUC18, pUC19, pUCI18, pUC119, and Bluescript M13, all of which are described in Sections 1.12-1.20 of Sambrook et al., *supra*. However, many other suitable vectors are available as well. These vectors contain genes coding for ampicillin and/or tetracycline resistance which enables cells transformed with these vectors to grow in the presence of these antibiotics.

The promoters most commonly used in prokaryotic vectors include the β-lactamase (penicillinase) and lactose promoter systems (Chang et al. *Nature* 375:615 (1978); Itakura et al., *Science* 198:1056 (1977); Goeddel et al., *Nature* 281:544 (1979)) and a tryptophan (trp) promoter system (Goeddel et al., *Nucl. Acids Res.* 8:4057 (1980); EPO Appl. Publ. No. 36,776), and the alkaline phosphatase systems. While these are the most commonly used, other microbial promoters have been utilized, and details concerning their nucleotide sequences have been published, enabling a skilled worker to ligate them functionally into plasmid vectors (see Siebenlist et al., *Cell* 20:269 (1980)).

Many eukaryotic proteins normally secreted from the cell contain an endogenous secretion signal sequence as part of the amino acid sequence. Thus, proteins normally found in the cytoplasm can be targeted for secretion by linking a signal sequence to the protein. This is readily accomplished by ligating DNA encoding a signal sequence to the 5' end of the DNA encoding the protein and then expressing this fusion protein in an appropriate host cell. The DNA encoding the signal sequence may be obtained as a restriction fragment from any gene encoding a protein with a signal sequence. Thus, prokaryotic, yeast, and eukaryotic signal sequences may be used herein, depending on the type of host cell utilized to practice the invention. The DNA and amino acid sequence encoding the signal sequence

portion of several eukaryotic genes including, for example, human growth hormone, proinsulin, and proalbumin are known (see Stryer, *Biochemistry* W.H. Freeman and Company, New York, NY, p. 769 (1988)), and can be used as signal sequences in appropriate eukaryotic host cells. Yeast signal sequences, as for example acid phosphatase (Arima et al., *Nucleic Acids Res.* 11:1657 (1983)), alpha-factor, alkaline phosphatase and invertase may be used to direct secretion from yeast host cells. Prokaryotic signal sequences from genes encoding, for example, LamB or OmpF (Wong et al., *Gene* 68:193 (1988)), MalE, PhoA, or beta-lactamase, as well as other genes, may be used to target proteins from prokaryotic cells into the culture medium.

Trafficking sequences from plants, animals and microbes can be employed in the practice of the invention to direct the gene product to the cytoplasm, endoplasmic reticulum, mitochondria or other cellular components, or to target the protein for export to the medium. These considerations apply to the overexpression of pinoresinol/lariciresinol reductase or dirigent protein, and to direction of expression within cells or intact organisms to permit gene product function in any desired location.

The construction of suitable vectors containing DNA encoding replication sequences, regulatory sequences, phenotypic selection genes and the dirigent protein DNA or pinoresinol/lariciresinol reductase DNA of interest are prepared using standard recombinant DNA procedures. Isolated plasmids and DNA fragments are cleaved, tailored, and ligated together in a specific order to generate the desired vectors, as is well known in the art (see, for example, Sambrook et al., *supra*).

As discussed above, pinoresinol/lariciresinol reductase variants, or dirigent protein variants, are preferably produced by means of mutation(s) that are generated using the method of site-specific mutagenesis. This method requires the synthesis and use of specific oligonucleotides that encode both the sequence of the desired mutation and a sufficient number of adjacent nucleotides to allow the oligonucleotide to stably hybridize to the DNA template.

A dirigent protein gene and/or pinoresinol/lariciresinol reductase gene, or an antisense nucleic acid fragment complementary to all or part of a dirigent protein gene or pinoresinol/lariciresinol reductase gene, may be introduced, as appropriate, into any plant species for a variety of purposes including, but not limited to: altering or improving the color, texture, durability and pest-resistance of wood tissue, especially heartwood tissue; reducing the formation of lignans and/or lignins in plant species, such as corn, which are useful as animal fodder, thereby enhancing the

availability of the cellulose fraction of the plant material to the digestive system of animals ingesting the plant material; reducing the lignan/lignin content of plant species utilized in pulp and paper production, thereby making pulp and paper production easier and cheaper; improving the defensive capability of a plant against predators and pathogens by enhancing the production of defensive lignans or lignins; the alteration of other ecological interactions mediated by lignans or lignins; producing elevated levels of optically-pure lignan enantiomers as medicines or food additives; introducing, enhancing or inhibiting the production of dirigent proteins or pinoresinol/lariciresinol reductases, or the production of pinoresinol or lariciresinol and their derivatives. A dirigent protein and/or pinoresinol/lariciresinol reductase gene may be introduced into any organism for a variety of purposes including, but not limited to: introducing, enhancing or inhibiting the production of dirigent protein and/or pinoresinol/lariciresinol reductase, or the production of pinoresinol or lariciresinol and their derivatives.

The foregoing may be more fully understood in connection with the following representative examples, in which "Plasmids" are designated by a lower case p followed by an alphanumeric designation. The starting plasmids used in this invention are either commercially available, publicly available on an unrestricted basis, or can be constructed from such available plasmids using published procedures. In addition, other equivalent plasmids are known in the art and will be apparent to the ordinary artisan.

"Digestion", "cutting" or "cleaving" of DNA refers to catalytic cleavage of the DNA with an enzyme that acts only at particular locations in the DNA. These enzymes are called restriction endonucleases, and the site along the DNA sequence where each enzyme cleaves is called a restriction site. The restriction enzymes used in this invention are commercially available and are used according to the instructions supplied by the manufacturers. (See also Sections 1.60-1.61 and Sections 3.38-3.39 of Sambrook et al., *supra*.)

"Recovery" or "isolation" of a given fragment of DNA from a restriction digest means separation of the resulting DNA fragment on a polyacrylamide or an agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. This procedure is known generally. For example, see Lawn et al.

(*Nucleic Acids Res.* 9:6103-6114 (1982)), and Goeddel et al. (*Nucleic Acids Res.*, *supra*).

The following examples merely illustrate the best mode now contemplated for practicing the invention, but should not be construed to limit the invention. All literature citations herein are expressly incorporated by reference.

EXAMPLE 1

Purification of Dirigent Protein from *Forsythia intermedia*

Plant Materials. Forsythia intermedia plants were either obtained from Bailey's Nursery (var. Lynwood Gold, St., Paul, MN), and maintained in Washington State University greenhouse facilities, or were gifts from the local community.

Initial Extraction and Ammonium Sulphate Precipitation. Solubilization of bound proteins was carried out at 4°C. Frozen *Forsythia intermedia* stems (2 kg) were pulverized in a Waring Blendor (Model CB6) in the presence of liquid nitrogen. The resulting powder was homogenized with 0.1 M KH_2PO_4-K_2HPO_4 buffer (pH 7.0, 4 liters) containing 5 mM dithiothreitol, and filtered through four layers of cheesecloth. The insoluble residue was consecutively extracted, with continuous agitation at 250 rpm, as follows: with chilled (-20°C) re-distilled acetone (4 liters, 3 x 30 min); 0.1 M KH_2PO_4-K_2HPO_4 buffer (pH 6.5) containing 0.1% β-mercaptoethanol (solution A, 8 liters, 30 min); solution A containing 1% Triton X100 (8 liters, 4 hours) and finally solution A (8 liters, 16 hours). Between each extraction, the residue was filtered through one layer of Miracloth (Calbiochem). Solubilization of the (+)-pinoresinol forming system was achieved by mechanically stirring the residue in solution A containing 1 M NaCl (8 liters, 4 hours). The homogenate was decanted and the resulting solution consecutively filtered through Miracloth (Calbiochem) and glass fiber (G6, Fisher Sci.). The filtrate was concentrated in an Amicon cell (Model 2000, YM 30 membrane) to a final volume of ~800 ml, and subjected to $(NH_4)_2SO_4$ fractionation. Proteins precipitating between 40 and 80% saturation were recovered by centrifugation (15,000*g*, 30 min) and the $(NH_4)_2SO_4$ pellet stored at -20°C until required.

Mono S Column Chromatography. Purification of 78-kD dirigent protein and partial purification of oxidase. The ammonium sulfate pellet (obtained from 2 kg of *F intermedia* stems) was reconstituted in 40 mM MES [2-(N-Morpholino)ethanesulfonic acid] buffer, adjusted to pH 5.0 with 6 M NaOH (solution B, 30 ml), the slurry being centrifuged (3,600*g*, 5 min), and the supernatant dialyzed overnight against solution B (4 liters). The dialyzed extract was filtered

(0.22 µm) and the sample (35 to 40 mg proteins) was applied to a MonoS HR5/5 (50 mm by 5 mm) column equilibrated in solution B at 4°C. After eluting (flow rate 5 ml min^{-1} cm^{-2}) with solution B (13 ml), proteins were desorbed with Na_2SO_4 in solution B, using a linear gradient from 0 to 100 mM in 8 ml and holding at this concentration for 32 ml, then implementing a series of step gradients at 133 mM for 50 ml, 166 mM for 50 ml, 200 mM for 40 ml, 233 mM for 40 ml and finally 333 mM Na_2SO_4 for 40 ml. Fractions capable of forming (+)-pinoresinol from *E*-coniferyl alcohol were eluted with 333 mM Na_2SO_4, combined and stored (-80°C) until needed.

POROS SP-M Matrix Column Chromatography (First Column). Fractions from 15 individual elutions from the MonoS HR5/5 column (33mM Na_2SO_4) were combined (18.5 mg proteins, 180 ml) and dialyzed overnight against solution C. The dialyzed enzyme solution (190 ml) was filtered (0.22 µm) and an aliquot (47 ml) was applied to the POROS SP-M column. All separations on a POROS SP-M matrix (100 mm by 4.6 mm), previously equilibrated in 25 mM MES-HEPES-sodium acetate buffer (pH 5.0, solution C), were performed at a flow rate of 60 ml min^{-1} cm^{-2} and at room temperature. After elution with solution C (12 ml), the proteins were desorbed with a linear Na_2SO_4 gradient (0 to 0.7 M in 66.5 ml) in solution C, whereupon the concentration established was held for an additional 16.6 ml. Under these conditions, separation of four fractions (I, II, III and IV) was achieved at ~40, 47, 55 and 61 mS, respectively. This purification step was repeated three times with the remaining dialyzed enzymatic extract, and fractions I, II, III, and IV from each experiment were separately combined. When protease inhibitors [that is, phenyl-methanesulfonyl fluoride (0.1 mmol ml^{-1}), EDTA (0.5 nmol ml^{-1}), pepstatin A (1 µg ml^{-1}), and antipain (1 µg ml^{-1})] were added during the solubilization and all subsequent purification stages, no differences were observed in the elution profiles of fractions I, II, III, and IV.

POROS SP-M Matrix Column Chromatography (Second Column). Fraction I from the first *POROS SP-M Matrix column chromatography* step (2.62 mg proteins, 40 ml, ~24.6 mS) was diluted in filtered, cold distilled water until the conductivity reached ~8 mS (final volume = 150 ml). The diluted protein solution was then applied onto a POROS SP-M column (100 mm by 4.6 mm). After elution with solution C (12 ml), fraction I was desorbed using a linear Na_2SO_4 gradient from 0 to 0.25 M in 20 ml, whereupon the concentration established was held for another 25 ml. This was followed by another linear Na_2SO_4 gradient from 0.25 to 0.7 M in 26 ml which was then held at 0.7 M for an additional 16.6 ml. Fractions eluted at ~30 mS (the ionic strength of the eluent was measured with a flow-through detector) were combined (15 ml, 1.3 mg), diluted with water and rechromatographed. The resulting protein (eluted at ~30 mS with the gradient described above) was stored (-80°C) until needed.

Gel filtration. An aliquot from fraction I (595.5 µg proteins, 3 ml, eluted at ~30 mS), was concentrated to 0.6 ml (Centricon 10, Amicon) and loaded onto a S200 (73.2 cm by 1.6 cm, Pharmacia-LKB) gel chromatographic column equilibrated in 0.1 M MES-HEPES-sodium acetate buffer (pH 5.0) containing 50 mM Na_2SO_4 at 4°C. An apparently homogenous 78-kD dirigent protein (242 µg) was eluted (flow rate 0.25 ml min^{-1} cm^{-2}) as a single component at 133 ml (Vo = 105 ml). Molecular weights were estimated by comparison of their elution profiles with the standard proteins, ß-amylase (200,000), alcohol dehydrogenase (150,000), bovine serum albumin (66,000), ovalbumin (45,000), carbonic anhydrase (29,000) and cytochrome c (12,400).

EXAMPLE 2

Characterization of the Purified Dirigent Protein

Molecular Weight and Isoelectric Point Determination. Polyacrylamide gel electrophoresis (PAGE) was performed in Laemmli's buffer system with gradient (4 to 15% acrylamide, Bio-Rad) gels under denaturing and reducing conditions. Proteins were visualized by silver staining. Gel filtration (S200) chromatography of fraction I gave a protein of native molecular weight ~78 kD, whereas SDS-polyacrylamide gel electrophoresis showed a single band at ~27 kD, suggesting that the native protein exists as a trimer. Isoelectric focusing of the native protein on a polyacrylamide gel (pH 3 to 10 gradient) revealed the presence of six bands. After isoelectric focusing, each of these bands was electroblotted onto a polyvinylidene fluoride (PVDF) membrane and subjected to amino terminal sequencing, which established that all had similar sequences indicating a series of isoforms. The ultraviolet-visible spectrum of the protein had only a characteristic protein absorbance at 280 nm with a barely perceptible shoulder at ~330 nm. Inductively coupled plasma (ICP) analysis gave no indication of any metal being present in the protein. Thus, the 78-kD dirigent protein lacks any detectable catalytically active oxidative center.

Assay of the Ability of the Purified Dirigent Protein to Form (+)Pinoresinol from E-Coniferyl alcohol. The four fractions (I to IV) from the first POROS SP-M

chromatographic step (Example 1) were individually rechromatographed, with each fraction subsequently assayed for (+)-pinoresinol-forming activity with E-[9-^{3}H]coniferyl alcohol as substrate for one hour. Fraction I (containing dirigent protein) had very little (+)-pinoresinol-forming activity (<5% of total activity loaded onto the POROS SP-M column), whereas fraction III catalyzed nonspecific oxidative coupling to give (±)-dehydrodiconiferyl alcohols, (±)-pinoresinols, and (±)-erythro/threo guaiacylglycerol 8-*O*-4'-coniferyl alcohol ethers. Thus, Fraction III appeared to contain an endogenous plant oxygenating protein.

Although the putative oxidase preparation (Fraction III) was not purified to electrophoretic homogeneity, the electron paramagnetic resonance (EPR) spectrum of this protein preparation resembled that of a typical plant laccase, *i.e.*, a class of naturally-occurring plant oxygenase proteins. We then studied the fate of E-[9-^{3}H]coniferyl alcohol (2 μmol ml^{-1}, 14.7 kBq) in the presence of, respectively, the oxidase (fraction III), the 78-kD dirigent protein (Fraction I), and both fraction III and the 78-kD protein together. With the fraction III preparation alone, only nonspecific bimolecular radical coupling occurs to give (±)-dehydrodiconiferyl alcohols, (±)-pinoresinols and (±)-erythro/threo guaiacylglycerol 8-*O*-4'-coniferyl alcohol ethers. With the 78-kD protein by itself, however, a small amount of (+)-pinoresinol formation (<5% over 10 hours) was observed, this being presumed to result from residual traces of oxidizing capacity in the preparation. When both fraction III and the 78-kD protein were combined, full catalytic activity and regio- and stereo-specificity in the product was reestablished, whereby essentially only (+)-pinoresinol was formed. Additionally, with fraction III alone, and when fraction III was combined with the 78-kD protein, the rates of substrate depletion and dimeric product formation were nearly identical. Moreover, essentially no turnover of the dimeric lignan products occurred in either case in the presence of the oxidase, over the time-period (8 hours) examined: subsequent dimer oxidation does not occur when *E*-coniferyl alcohol, the preferred substrate, is still present in the assay mixture. The 78-kD protein therefore appears to determine the specificity of the bimolecular phenoxy radical coupling reaction.

Gel filtration studies were also carried out with mixtures of the dirigent and fraction III proteins, in order to establish if any detectable protein-protein interaction might account for the stereoselectivity. But no evidence in support of complex formation (i.e., to higher molecular size entities) was observed.

EXAMPLE 3

Effect of the 78-KD Dirigent Protein on Plant Laccase-Catalyzed Monolignol Coupling

E-coniferyl alcohol coupling assay. E-[9-^{3}H]Coniferyl alcohol (4 μmol ml^{-1}, 29.3 kBq) was incubated with a 120-kD laccase (previously purified from *Forsythia intermedia* stem tissue) over a 24-hour period, in the presence and absence of the dirigent protein, as follows. Each assay consisted of E-[9-^{3}H]coniferyl alcohol (4 μmol ml^{-1}, 29.3 kBq, 7.3 MBq mole liter^{-1}; or 2 μmol ml^{-1}, 14.7 kBq with fraction III), the 78-kD dirigent protein, an oxidase or oxidant, or both [final concentrations: 770 pmol ml^{-1} dirigent protein; 10.7 pmol protein ml^{-1} *Forsythia* laccase; 12 μg protein ml^{-1} fraction III; 0.5 μmol ml^{-1} FMN; 0.5 μmol ml^{-1} FAD; 1 and 10 μmol ml^{-1} ammonium peroxydisulfate] in buffer (0.1 M MES-HEPES-sodium acetate, pH 5.0) to a total volume of 250 μl. The enzymatic reaction was initiated by addition of E-[9-^{3}H]coniferyl alcohol. Controls were performed in the presence of buffer alone.

After one hour incubation at 30 °C while shaking, the assay mixture was extracted with ethyl acetate (EtOAc, 500 μl) containing (±)-pinoresinols (7.5 μg), (±)-dehydrodiconiferyl alcohols (3.5 μg) and erythro/threo (±)-guaiacylglycerol 8-*O*-4'-coniferyl alcohol ethers (7.5 μg) as radiochemical carriers and ferulic acid (15.0 μg) as an internal standard. After centrifugation (13,800*g*, 5 min), the EtOAc soluble components were removed and the extraction procedure repeated with EtOAc (500 μl). The EtOAc soluble components from each assay were combined, the solutions evaporated to dryness *in vacuo*, redissolved in methanol-water solution (1:1; 100 μl) with an aliquot (50 μl) thereof subjected to reversed-phase column chromatography (Waters, Nova-Pak C_{18}, 150 mm by 3.8 mm). The elution conditions were as follows: acetonitrile/3 % acetic acid in H_2O (5:95) from 0 to 5 min, then linear gradients to ratios of 10:90 between 5 and 20 min, then to 20:80 between 20 and 45 min and finally to 50:50 between 45 and 60 min, at a flow rate of 8.8 ml min^{-1} cm^{-2}.

Fractions corresponding to *E*-coniferyl alcohol, erythro/threo (±)-guaiacylglycerol 8-*O*-4'-coniferyl alcohol ethers, (±)-dehydrodiconiferyl alcohols and (±)-pinoresinols were individually collected, aliquots removed for liquid scintillation counting, and the remainder freeze-dried. Pinoresinol-containing fractions were redissolved in methanol (100 μl) and subjected to chiral column chromatography (Daicel, Chiralcel OD, 50 mm by 4.6 mm) with a solution of

hexanes and ethanol (1:1) as the mobile phase (flow rate 3 ml min^{-1} cm^{-2}), whereas dehydrodiconiferyl alcohol fractions were subjected to Chiralcel OF (250 mm by 4.6 mm) column chromatography eluted with a solution of hexanes and isopropanol (9:1) as the mobile phase (flow rate 2.4 ml min^{-1} cm^{-2}), the radioactivity of the eluent being measured with a flow-through detector (Radiomatic, Model A120).

Results of E-coniferyl alcohol coupling assay. Incubation with laccase alone gave only racemic dimeric products, with (±)-dehydrodiconiferyl alcohols predominating. In the presence of the dirigent protein, however, the process was now primarily stereoselective, affording (+)-pinoresinol, rather than being nonspecific as observed when only laccase was present. The rates of both *E*-coniferyl alcohol (substrate) depletion and the formation of the dimeric lignans were similar with and without the dirigent protein. A substantial difference was noted in the subsequent turnover of the lignan products observed after *E*-coniferyl alcohol depletion. With the laccase alone no turnover occurred, but when both proteins were present the disappearance of the products was significant. In order to understand the difference, assays were conducted where bovine serum albumin (BSA) and ovalbumin were individually added to the laccase-containing solutions at levels matching the weight concentrations of the dirigent protein. In this way, it was established that the differences in product turnover were simply due to stabilization of laccase activity at the higher protein concentrations, although interestingly the dirigent protein, BSA and ovalbumin afforded somewhat different degrees of protection. The findings were quite comparable when a fungal laccase (from *Trametes versicolor*) was used in place of the plant laccase. When the oxidizing capacity (i.e., laccase concentration) was lowered five-fold, only (+)-pinoresinol formation was observed. Thus, complete stereoselectivity is preserved when the oxidative capacity does not exceed a point where the dirigent protein is saturated.

Stereoselective E-coniferyl alcohol coupling. Assays were also conducted with *E*-[9-$^2H_2$, OC^2H_3]coniferyl alcohol and the dirigent protein in the presence of laccase as follows. *E*-[9-$^2H_2$, OC^2H_3]coniferyl alcohol (2 μmol ml^{-1}) was incubated in the presence of dirigent protein (770 pmol ml^{-1}), the purified plant laccase (4.1 pmol ml^{-1}) and buffer (0.1 M MES-HEPES-sodium acetate, pH 5.0) in a total volume of 250 μl. After one hour incubation, the reaction mixture was extracted with EtOAc, but with the addition of an internal standard and radiochemical carriers omitted. After reversed-phase column chromatography, the enzymatically formed pinoresinol was collected, freeze-dried, redissolved in methanol (100 μl) and

subjected to chiral column chromatography (Daicel, Chiralcel OD, 50 mm by 4.6 mm) with detection at 280 nm and analysis by mass spectral fragmentation in the EI mode (Waters, Integrity System). Liquid chromatography-mass spectrometry (LC-MS) analysis of the resulting (+)-pinoresinol (>99% enantiomeric excess) gave a molecular ion with a mass to charge ratio (m/z) 368, thus establishing the presence of 10 ^2H atoms and verifying that together the laccase and dirigent protein catalyzed stereoselective coupling of *E*-[9-$^2H_2$, OC^2H_3]coniferyl alcohol.

Other auxiliary one-electron oxidants can also facilitate stereoselective coupling with the dirigent protein. Ammonium peroxydisulfate readily undergoes homolytic cleavage (A. Usaitis, R. Makuska, *Polymer* 35:4896 (1994)) and is routinely used as an one-electron oxidant in acrylamide polymerization. Ammonium peroxydisulfate was first incubated with *E*-[9-^3H]coniferyl alcohol (4 μmol ml^{-1}, 29.3 kBq) for 6 hours using the *E*-coniferyl alcohol coupling assay procedure described above. Nonspecific bimolecular radical coupling was observed, to afford predominantly (±)-dehydrodiconiferyl alcohols as well as the other racemic lignans (Table 1). However, when the dirigent protein was added, the stereoselectivity of coupling was dramatically altered to give primarily (+)-pinoresinol at both concentrations of oxidant, together with small amounts of racemic lignans. This established that even an inorganic oxidant, such as ammonium peroxydisulfate, could promote (+)-pinoresinol synthesis in the presence of the dirigent protein, even if it was not oxidatively as selective toward the monolignol as was the fraction III oxidase or laccase.

Table 1.
Effect of dirigent protein on product distribution from *E*-coniferyl alcohol oxidized by ammonium peroxydisulfate (6 hour assay).

Oxidant	Dirigent protein (770 pmol ml^{-1})	*E*-Coniferyl alcohol in dimer equivalents depleted (nmol ml^{-1})	(±)-Guaiacyl-glycerol 8-*O*-4'-coniferyl alcohol ethers (nmol ml^{-1})	(±)-Dehydro-diconiferyl alcohols (nmol ml^{-1})	(±)-Pinoresinols (nmol ml^{-1})	(+)-Pinoresinol (nmol ml^{-1})	Total dimers (nmol ml^{-1})
Ammonium peroxydisulfate	absent	200 ± 4	10 ± 1	35 ± 2	16 ± 0	0	61 ± 3
(1 μmol ml^{-1})	present	250 ± 55	6 ± 0	13 ± 1	0	130 ± 10	149 ± 11
Ammonium peroxydisulfate	absent	860 ± 30	90 ± 4	250 ± 10	135 ± 4	0	475 ± 17
(10 μmol ml^{-1})	present	1030 ± 25	30 ± 1	90 ± 3	0	450 ± 10	570 ± 14
Dirigent protein	present	61 ± 20	5 ± 1	8 ± 1	0	55 ± 1	68 ± 3

Effect of Other Oxygenating Agents on the Stereospecific Conversion of E-Coniferyl Alcohol to (+)-pinoresinol. The effects of incubating *E*-coniferyl alcohol (4 μmol ml^{-1}, 29.3 kBq) with flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) were investigated since, in addition to their roles as enzyme cofactors, they can also oxidize various organic substrates (T.C. Bruice, *Acc. Chem. Res.* 13:256 (1980)). *E*-[9-^{3}H]coniferyl alcohol was respectively incubated with FMN and FAD for 48 hours. To obtain the FMN, snake (*Naja naja atra*, Formosan cobra) venom was added to a solution of FAD (5 μmol ml^{-1} in H_2O) and, after 30 min incubation at 30°C, the enzymatically formed FMN was separated from the protein mixture by filtration through a Centricon 10 (Amicon) microconcentrator. In every instance, *E*-coniferyl alcohol oxidation was more rapid in the presence of FMN than FAD. Although these differences between the FMN and FAD catalyzed rates of *E*-coniferyl alcohol oxidation were not anticipated, a consistent pattern was sustained: racemic lignan products were obtained, with the (±)-dehydrodiconiferyl alcohols predominating as before. When the time courses were repeated in the presence of the dirigent protein, a dramatic change in stereoselectivity was observed, where essentially only (+)-pinoresinol formation occurred. Again, the rates of *E*-coniferyl alcohol depletion, when adjusted for the traces of residual oxidizing capacity (<5% over 10 hours) in the dirigent protein preparation, were dependent only upon [FMN] and [FAD], as were the total amounts of dimers formed. When full depletion of *E*-coniferyl alcohol occurs, the corresponding lignan dimers can begin to undergo oxidative changes as a function of time; specifically, FMN is able subsequently to oxidize pinoresinol, in open solution, after the *E*-coniferyl alcohol has been fully depleted.

Investigation of Substrate-Specific Stereoselectivity. The coupling stereoselectivity was substrate specific. Neither *E-p*-[9-^{3}H]coumaryl (4 μmol ml^{-1}, 44.5 kBq) or *E*-[8-^{14}C]sinapyl alcohols (4 μmol ml^{-1}, 8.3 kBq), which differ from *E*-coniferyl alcohol only by a methoxyl group substituent on the aromatic ring, yielded stereoselective products when incubated for 6 hours with FMN and ammonium peroxydisulfate respectively, in the presence and absence of the dirigent protein. Incubations were carried out as described above with the following modifications: *E-p*-[9-^{3}H]coumaryl (4 μmol ml^{-1}, 44.5 kBq) or *E*-[8-^{14}C]sinapyl alcohols (4 μmol ml^{-1}, 8.3 kBq) were used as substrates and, after 6 hour incubation at 30°C, the reaction mixture was extracted with EtOAc but without addition of radiochemical carriers. *E*-Sinapyl alcohol readily underwent coupling to afford

syringaresinol, but chiral HPLC analysis revealed that the resulting products were, in every instance, racemic (Table 2). Interestingly, by itself, the 78-kD dirigent protein preparation catalyzed a low level of dimer formation, as previously noted, but only gave rise to racemic (±)-syringaresinol formation, which is presumably a consequence of the residual traces of contaminating oxidizing capacity present in the protein preparation.

In an analogous manner, no stereoselective coupling was observed with *E-p*-coumaryl alcohol as substrate. That is, only *E*-coniferyl alcohol undergoes stereoselective coupling in the presence of the dirigent protein. Given the marked substrate specificity of the dirigent protein for *E*-coniferyl alcohol, it will be of considerable interest to determine, in the future, how it differs from that affording (+)-syringaresinol in *Eucommia ulmoides* (T. Deyama, *Chem. Pharm. Bull.* 31, 2993 (1983)).

Table 2.

Effect of dirigent protein on coupling of *E*-sinapyl alcohol (6 hour assay).

	Dirigent protein (770 pmol ml^{-1})	*E*-Sinapyl alcohol in dimer equivalents depleted (nmol ml^{-1})	Racemic (±)-syringaresinols (nmol ml^{-1})
FMN	absent	570 ± 100	290 ± 40
(0.5 µmol ml^{-1})	present	610 ± 110	340 ± 40
Ammonium peroxydisulfate	absent	1400 ± 120	1020 ± 40
(10 µmol ml^{-1})	present	1520 ± 10	1060 ± 30
Dirigent protein	present	110 ± 10	50 ± 10

Although the inventors do not intend to be bound by any particular mechanism for stereoselective coupling, three distinct possibilities can be envisaged. The most likely is that the oxidase or oxidant generates free-radical species from *E*-coniferyl alcohol, and that the latter are the true substrates that bind to the dirigent protein prior to coupling. The other two possibilities would require that *E*-coniferyl alcohol molecules are bound and oriented on the dirigent protein, thereby ensuring that only (+)-pinoresinol formation occurs upon subsequent oxidative coupling: this

could occur either if both substrate phenolic hydroxyl groups were exposed so that they could readily be oxidized by an oxidase or oxidant, or if an electron transfer mechanism were operative between the oxidase or oxidant and an electron acceptor site or sites on the dirigent protein.

Among the three alternative mechanisms, three lines of evidence suggest "capture" of phenoxy radical intermediates by the dirigent protein. First, the rates of both substrate depletion and product formation are largely unaffected by the presence of the dirigent protein. If capture of the free-radical intermediates is the operative mechanism, the dirigent protein would only affect the specificity of coupling when single-electron oxidation of coniferyl alcohol is rate-determining. Second, an electron transfer mechanism is currently ruled out, since we observed no new ultraviolet-visible chromophores in either the presence or absence of an auxiliary oxidase or oxidant, under oxidizing conditions. Third, preliminary kinetic data (as disclosed in Example 4) support the concept of free-radical capture based on the formal values of Michaelis constant (K_m) and maximum velocity (V_{max}) characterizing the conversion of *E*-coniferyl alcohol into (+)-pinoresinol, with the dirigent protein alone and in the presence of the various oxidases or oxidants.

EXAMPLE 4

Kinetic Characterization of the Conversion of E-Coniferyl Alcohol to (+)-pinoresinol in the Presence of Dirigent Protein and an Oxygenating Agent.

Assays were carried out as described in Example 3 by incubating a series of *E*-[9-^{3}H]coniferyl alcohol concentrations (between 8.00 and 0.13 µmol ml^{-1}, 7.3 MBq mole liter^{-1}) with dirigent protein (770 pmol ml^{-1}) alone and in presence of *Forsythia* laccase (2.1 pmol ml^{-1}), fraction III (12 µg protein ml^{-1}), or FMN (0.5 µmol ml^{-1}). Assays with dirigent protein, in presence or absence of FMN, were incubated at 30°C for 1 hour, whereas assays with *Forsythia* laccase or fraction III in presence or absence of dirigent protein were incubated at 30 °C for 15 min. If free-radical capture by the dirigent protein is the operative mechanism, the Michaelis-Menten parameters obtained will only represent formal rather than true values, because the highest free-energy intermediate state during the conversion of *E*-coniferyl alcohol into (+)-pinoresinol is still unknown and the relation between the concentration of substrate and that of the corresponding intermediate free-radical in open solution has not been delineated.

Bearing these qualifications in mind, we estimated formal K_m and V_{max} values for the dirigent protein preparation. As noted earlier, it was capable of

engendering formation of low levels of both (+)-pinoresinol from *E*-coniferyl alcohol, and racemic (±)-syringaresinols from *E*-sinapyl alcohol, because of traces of contaminating oxidizing capacity. With this preparation (Table 3), a formal K_m of 10 ± 6 mM and V_{max} of 0.02 ± 0.02 mol s^{-1} mol^{-1} were obtained. However, with addition of fraction III, laccase, and FMN, the formal K_m values (mM) were reduced to 1.6 ± 0.3, 0.100 ± 0.003, and 0.10 ± 0.01, respectively, whereas the V_{max} values were far less affected at these concentrations of auxiliary oxidase/oxidant.

Formal K_m and V_{max} values were calculated for the laccase and fraction III oxidase with respect to *E*-coniferyl alcohol conversion into the three racemic lignans. However, no direct comparisons can be made to the 78-kD protein, since the formal K_m values involve only the corresponding oxidases. For completeness, the K_m (mM) and V_{max} (mol s^{-1} mol^{-1} enzyme) were as follows: with respect to the laccase, 0.200 ± 0.001 and 3.9 ± 0.2 for (±)-erythro/threo guaiacylglycerol 8-*O*-4'-coniferyl alcohol ethers, 0.3000 ± 0.0003 and 13.1 ± 0.6 for (±)-dehydrodiconiferyl alcohols, and 0.300 ± 0.002 and 7.54 ± 0.50 for (±)-pinoresinols; with respect to the fraction III oxidase (estimated to have a native molecular weight of 80 kDa), 2.2 ± 0.3 and 0.20 ± 0.03 for (±)-erythro/threo guaiacylglycerol 8-*O*-4'- coniferyl alcohol ethers, 2.2 ± 0.2 and 0.7 ± 0.1 for (±)-dehydrodiconiferyl alcohols, and 3.7 ± 0.7 and 0.6 ± 0.1 for (±)-pinoresinols.

These preliminary kinetic parameters are in harmony with the finding that dirigent protein does not substantially affect the rate of *E*-coniferyl alcohol depletion in the presence of fraction III, laccase and FMN. Both sets of results are together in accord with the working hypothesis that the dirigent protein functions by capturing free-radical intermediates which then undergo stereoselective coupling.

Table 3.
Effect of various oxidants on formal K_m and V_{max} values for the dirigent protein (770 pmol ml^{-1}) during (+)-pinoresinol formation from *E*-coniferyl alcohol.

Oxidase/Oxidant	Formal K_m (mM)	V_{max} (mol s^{-1} mol^{-1} dirigent protein)
Dirigent protein	10 ± 6	0.02 ± 0.02
Fraction III (12 μg protein ml^{-1})	1.6 ± 0.3	0.10 ± 0.03
Laccase (2.07 pmol ml^{-1})	0.100 ± 0.003	0.0600 ± 0.0002
FMN (0.5 μmol ml^{-1})	0.10 ± 0.01	0.024 ± 0.001

EXAMPLE 5

Cloning of the Dirigent Protein cDNA From *Forsythia intermedia*

Plant Materials - *Forsythia intermedia* plants were either obtained from Bailey's Nursery (var. Lynwood Gold, St., Paul, MN), and maintained in Washington State University greenhouse facilities, or were gifts from the local community.

Materials - All solvents and chemicals used were reagent or HPLC grade. *Taq* thermostable DNA polymerase was obtained from Promega, whereas restriction enzymes were from Gibco BRL (*Hae*III), Boehringer Mannheim (*Sau*3a) and Promega (*Taq*I). pT7Blue T-vector and competent NovaBlue cells were purchased from Novagen and radiolabeled nucleotide ([α-^{32}P]dCTP) was from DuPont NEN.

Oligonucleotide primers for polymerase chain reaction (PCR) and sequencing were synthesized by Gibco BRL Life Technologies. GENECLEAN II® kits (BIO 101 Inc.) were used for purification of PCR fragments, with the gel-purified DNA concentrations determined by comparison to a low DNA mass ladder (Gibco BRL) in 1.5% agarose gels.

Instrumentation - UV (including RNA and DNA determinations at OD_{260}) spectra were recorded on a Lambda 6 UV/VIS spectrophotometer. A Temptronic II thermocycler (Thermolyne) was used for all PCR amplifications. Purification of DNA for sequencing employed a QIAwell Plus plasmid purification system (QIAGEN) followed by PEG precipitation (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1994) *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), with DNA sequences determined using an Applied Biosystems Model 373A automated sequencer. Amino acid sequences were obtained using an Applied Biosystems protein sequencer with on-line HPLC detection, according to the manufacturer's instructions.

Dirigent Protein Amino Acid Sequencing - The dirigent protein N-terminal amino acid sequence (SEQ ID No:1) was obtained from the purified protein using an Applied Biosystems protein sequencer with on-line HPLC detection. For trypsin digestion, the purified enzyme (150 pmol) was suspended in 0.1 M Tris-HCl (50 μl, pH 8.5, Boehringer Mannheim, sequencing grade), with urea added to give a final concentration of 8 M in 77.5 μl. The mixture was incubated for 15 min at 50°C, following which 100 mM iodoacetamide (2.5 μl) was added, with the whole kept at room temperature for 15 min. Trypsin (1 μg in 20 μl) was then added, with the mixture digested for 24 h at 37°C, following which TFA (4 μl) was added to stop the enzymatic reaction. The resulting mixture was subjected to reversed phase HPLC

analysis (C-8 column, Applied Biosytems), this being eluted with a linear gradient over 2 h from 0 to 100% acetonitrile (in 0.1% TFA) at a flow rate of 0.2 ml/min with detection at 280 nm. Fractions containing individual oligopeptide peaks were collected manually and directly submitted to amino acid sequencing (SEQ ID Nos:2-7).

Forsythia intermedia stem cDNA Library Synthesis - Total RNA (~300 μg/g fresh weight) was obtained (Dong, Z.D., and Dunstan, D.I. (1996) *Plant Cell Reports* 15:516-521) from young green stems of greenhouse-grown *Forsythia intermedia* plants (var. Lynwood Gold). A *Forsythia intermedia* stem cDNA library was constructed using 5 μg of purified poly A^+ mRNA (Oligotex-dT™ Suspension, QIAGEN) with the ZAP-cDNA® synthesis kit, the Uni-ZAP™ XR vector and the Gigapack® II Gold packaging extract (Stratagene), with a titer of 1.2×10^6 PFU for the primary library. A portion (30 ml) of the amplified library (1.2×10^{10} PFU/ml; 158 ml total) (Sambrook, J. et al., *supra*) was used to obtain pure cDNA library DNA (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidnam, J.G., Smith, J.A., and Struhl, K. (1991) *Current Protocols in Molecular Biology*, 2 volumes, Greene Publishing Associates and Wiley-Interscience, John Wiley & Sons, NY) for PCR.

Dirigent Protein DNA Probe Synthesis - The N-terminal and internal peptide amino acid sequences were used to construct the degenerate oligonucleotide primers. Purified *F. intermedia* cDNA library DNA (5 ng) was used as the template in 100 μl PCR reactions (10 mM Tris-HCl [pH 9.0], 50 mM KCl, 0.1% Triton X-100, 2.5 mM $MgCl_2$, 0.2 mM each dNTP and 2.5 units *Taq* DNA polymerase) with primer PSINT1 (SEQ ID No:8) (100 pmol) and either primer PSI7R (SEQ ID No:11) (20 pmol), primer PSI2R (SEQ ID No:10) (20 pmol) or primer PSI1R (SEQ ID No:9) (20 pmol). PCR amplification was carried out in a thermocycler as follows: 35 cycles of 1 min at 94°C, 2 min at 50°C and 3 min at 72°C; with 5 min at 72°C and an indefinite hold at 4°C after the final cycle. Single-primer, template-only and primer-only reactions were performed as controls. PCR products were resolved in 1.5% agarose gels, where a single band (~370-, ~155- or ~125-bp, respectively) was observed for each reaction.

To determine the nucleotide sequence of the amplified bands, five 100 μl PCR reactions were performed as above with PSINT1 (SEQ ID No:8) +PSI7R (SEQ ID No:11), PSINT1 (SEQ ID No:8) +PSI2R (SEQ ID No:10) and PSINT1 (SEQ ID No:8) +PSI1R (SEQ ID No:9) primer pairs. The 5 reactions from each primer pair were concentrated (Microcon 30, Amicon Inc.) and washed with TE

buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA; 2 x 200 μl), with the PCR products subsequently recovered in TE buffer (2 x 50 μl). These were resolved in preparative 1.5% agarose gels. Each gel-purified PCR product (~0.2 pmol) was then ligated into the pT7Blue T-vector and transformed into competent NovaBlue cells, according to Novagen's instructions. Insert sizes were determined using the rapid boiling lysis and PCR technique (with R20mer and U19mer primers) according to the manufacturer's instructions. Restriction analyses were performed to determine whether all inserts from the reactions utilizing each of the foregoing primer pairs were the same, as follows: to 20 μl each of a 100 μl PCR reaction (insert of interest amplified with R20mer(SEQ ID No:74) and U19mer(SEQ ID No:75) primers) were added 4 units *Hae*III, 1.5 units *Sau*3a or 5 units *Taq*I restriction enzyme. Restriction digestions were allowed to proceed for 60 min at 37°C for *Hae*III and *Sau*3A and at 65°C for *Taq*I reactions. Restriction products were resolved in 1.5% agarose gels giving one restriction group for each insert tested. Five recombinant plasmids from PSINT1 (SEQ ID No:8) +PSI7R (SEQ ID No:11) (called pT7PSI1-pT7PSI5) and 2 recombinant plasmids from PSINT1 (SEQ ID No:8) +PSI2R (SEQ ID No:10) (called pT7PSI6 and pT7PSI7) PCR products were selected for DNA sequencing; all contained the same open reading frame (ORF) (SEQ ID No:69). The dirigent protein probe was next constructed as follows: five 100 μl PCR reactions were performed as above with 10 ng pT7PSI1 DNA (SEQ ID No:69) with primers PSINT1 (SEQ ID No:8) and PSI7R (SEQ ID No:11). Gel-purified pT7PSI1 insert (50 ng) was used with Pharmacia's T7QuickPrime® kit and [α-^{32}P]dCTP, according to kit instructions, to produce a radiolabeled probe (in 0.1 ml), which was purified over BioSpin 6 columns (Bio-Rad) and added to carrier DNA (0.5 mg/ml sheared salmon sperm DNA [Sigma], 0.9 ml).

Library Screening - 600,000 PFU of *F. intermedia* amplified cDNA library were plated for primary screening, according to Stratagene's instructions. Plaques were blotted onto Magna Nylon membrane circles (Micron Separations Inc.), which were then allowed to air dry. The membranes were placed between two layers of Whatman® 3MM Chr paper. cDNA library phage DNA was fixed to the membranes and denatured in one step by autoclaving for 2 min at 100°C with fast exhaust. The membranes were washed for 30 min at 37°C in 6X standard saline citrate (SSC) and 0.1% SDS and prehybridized for 5 h with gentle shaking at 57-58°C in preheated 6X SSC, 0.5% SDS and 5X Denhardt's reagent (hybridization solution, 300 ml) in a crystallization dish (190 x 75 mm). The [^{32}P]radiolabeled probe was denatured

(boiling, 10 min), quickly cooled (ice, 15 min) and added to a preheated fresh hybridization solution (60 ml, 58°C) in a crystallization dish (150 x 75 mm). The prehybridized membranes were next added to this dish, which was then covered with plastic wrap. Hybridization was performed for 18 h at 57-58°C with gentle shaking. The membranes were washed in 4X SSC and 0.5% SDS for 5 min at room temperature, transferred to 2X SSC and 0.5% SDS (at room temperature) and incubated at 57-58°C for 20 min with gentle shaking, wrapped with plastic wrap to prevent drying and finally exposed to Kodak X-OMAT AR film for 24 h at -80°C with intensifying screens. Twenty positive plaques were purified through two more rounds of screening with hybridization conditions as above.

In vivo Excision and Sequencing of Dirigent Protein cDNA-containing Phagemids - Purified cDNA clones were rescued from the phage following Stratagene's *in vivo* excision protocol. Both strands of several different cDNAs that coded for dirigent protein were completely sequenced using overlapping sequencing primers. Two distinct cDNAs were identified, called pPSD_Fi1(SEQ ID No:12) and pPSD_Fi2(SEQ ID No:14).

Sequence Analysis - DNA and amino acid sequence analyses were performed using the Unix-based GCG Wisconsin Package (Program Manual for the Wisconsin Package, Version 8, September 1994, *Genetics Computer Group*, 575 Science Drive, Madison, Wisconsin, USA 53711; Rice, P. (1996) *Program Manual for the EGCG Package*, Peter Rice, The Sanger Centre, Hinxton Hall, Cambridge, CB10 1Rq, England) and the ExPASy World Wide Web molecular biology server (Geneva University Hospital and University of Geneva, Geneva, Switzerland).

EXAMPLE 6

Expression of Functional Dirigent Protein in *Spodoptera frugiperda*

Attempts to express functional dirigent protein in *Escherichia coli* failed. Consequently, we expressed the dirigent protein in *Spodoptera frugiperda* utilizing a baculovirus expression system. The full-length 1.2 kb cDNA clone for the dirigent protein (PSD) in *F. intermedia*, containing both the 5' and 3' untranslated regions, was excised from the pBlueScript (Stratagene) derived plasmid pPSD_Fi1 (SEQ ID No:12) using the restriction endonucleases *Bam*H I and *Xho* I. This 1.2 kb fragment was directionally subcloned into these same restriction sites in the multiple cloning site of the baculovirus transfer vector pBlueBac4 (Invitrogen, San Diego, CA). This produced the 6.0 kb construct pBB4/PSD which generates a non-fusion dirigent protein with translation being initiated at the dirigent protein cDNA

start codon. This construct was then co-transfected with linearized Bac-N-Blue DNA (Invitrogen) into *Spodoptera frugiperda* Sf9 cells by the technique of cationic liposome mediated transfection to produce, by means of homologous recombination, the recombinant *Autographa californica* nuclear polyhedrosis viral (AcMNPV) DNA Bac-N-Blue dirigent protein (BB/PSD) which was purified from plaques according to procedures described by Invitrogen. The final recombinant AcMNPV-BB/PSD contains the PSD gene under the polyhedrin promoter control and the essential sequence needed for replication of the recombinant virus. To verify that the dirigent protein was successfully expressed in the insect cell culture, log phase Sf9 cells infected with the AcMNPV-PSD recombinant viral high titer stock were used to obtain heterologous protein production. Maximal dirigent protein yield occurred by 48-70 hours post-infection. As determined by SDS-PAGE and (+)-pinoresinol forming activity, the protein was found secreted into the medium and showed a molecular mass and activity which corresponded to the indigenous protein originally isolated from *Forsythia intermedia.*

EXAMPLE 7

Isolation of Dirigent Protein Clones from *Thuja plicata* and *Tsuga heterophylla*

The coding region of a *Forsythia* dirigent protein cDNA, psd-Fi1 (SEQ ID No:12), was used to screen cDNA libraries from *Thuja plicata* and *Tsuga heterophylla*. The conditions and methods were as disclosed in Example 5, except that hybridization was carried out at 45-50°C. Two dirigent protein cDNAs were isolated from *Tsuga heterophylla* (SEQ ID Nos:16, 18), and eight dirigent protein cDNAs were isolated from *Thuja plicata* (SEQ ID Nos:20, 22, 24, 26, 28, 30, 32, 34).

EXAMPLE 8

Purification of Pinoresinol/lariciresinol Reductases from Forsythia Intermedia

Plant Materials. Forsythia intermedia plants were either obtained from Bailey's Nursery (var. Lynwood Gold, St., Paul, MN), and maintained in Washington State University greenhouse facilities, or were gifts from the local community.

Materials. All solvents and chemicals used were reagent or HPLC grade. Unlabeled (±)-pinoresinols and (±)-lariciresinols were synthesized as described (Katayama, T. et al., *Phytochemistry* 32:581-591 (1993)). [4R-3H]NADPH was obtained as previously reported (Chu, A. et al., *J. Biol. Chem.* 268:27026-27033 (1993)) by modification of the procedure of Moran et al. (Moran, R.G. et al., *Anal. Biochem.* 138:196-204 (1984)), and [4R-2H]NADPH was prepared according to

Anderson and Lin (Anderson, J.A., and Lin B.K., *Phytochemistry* 32:811-812 (1993)). Yeast glucose-6-phosphate dehydrogenase (Type IX,22.32. mmol h^{-1} mg^{-1}) and yeast hexokinase (Type F300, 15.12 $mmol^{-1}$ mg^{-1}) were purchased from Sigma and dihydrofolate reductase (*Lactobacillus casei*, 33.48 mmol h^{-1} mg^{-1}) was obtained from Biopure Co. Affi-Gel Blue Gel (100-200 mesh) and Bio-Gel HT Hydroxyapatite were purchased from Bio-Rad, whereas Phenyl Sepharose CL-4B, MonoQ HR 5/5, MonoP HR 5/20, Superose 6, Superose 12, Superdex 75, PD-10 columns, molecular weight standards and Polybuffer 74 were obtained from Pharmacia LKB Biotechnology, Inc. Adenosine 2',5'-diphosphate Sepharose and Reactive Yellow 3 Agarose were from Sigma Chemical Co.

Instrumentation. ^{1}H Nuclear magnetic resonance spectra (300 and 500 MHz) were recorded on Bruker AMX300 and Varian VXR500S spectrometers, respectively, using $CDCl_3$ as solvent with chemical shifts (δ ppm) reported downfield from tetramethylsilane (internal standard). UV (including RNA and DNA determinations at OD_{260}) and mass spectra were obtained on Lambda 6 UV/VIS and VG 7070E (ionizing voltage 70 eV) spectrophotometers, respectively. High performance liquid chromatography was carried out using either reversed-phase (Waters, Nova-pak C18, 150 x 3.9 mm inner diameter) or chiral (Daicel, Chiralcel OD or Chiralcel OC, 240 x 4.6 mm inner diameter) columns, with detection at 280 nm (Chu, A. et al., *J. Biol. Chem.* 268:27026-27033 (1993)). Radioactive samples were analyzed in Ecolume (ICN) and measured using a liquid scintillation counter (Packard, Tricarb 2000 CA). Amino acid sequences were obtained using an Applied Biosystems protein sequencer with on-line HPLC detection, according to the manufacturer's instructions.

Enzyme Assays. Pinoresinol and lariciresinol reductase activities were assayed by monitoring the formation of (+)-[^{3}H]lariciresinol and (-)-[^{3}H]secoisolariciresinol (Chu, A. et al., *J. Biol. Chem.* 268:27026-27033 (1993)).

Briefly, each assay for pinoresinol reductase activity consisted of (±)-pinoresinols (5 mM in MeOH, 20 µl), the enzyme preparation at the corresponding stage of purity (100 µl), and buffer (20 mM Tris-HCl, pH 8.0, 110 µl). The enzymatic reaction was initiated by addition of [4R-^{3}H]NADPH (10 mM, 6.79 kBq/mmol in 20 µl of double-distilled H_2O). After 30 min incubation at 30°C with shaking, the assay mixture was extracted with EtOAc (500 µl) containing (±)-lariciresinols (20 µg) and (±)-secoisolariciresinols (20 µg) as radiochemical carriers. After centrifugation (13,800 x g, 5 min), the EtOAc solubles were removed

and the extraction procedure was repeated. For each assay, the EtOAc solubles were combined with an aliquot (100 µl) removed for determination of its radioactivity using liquid scintillation counting. The remainder of the combined EtOAc solubles was evaporated to dryness in vacuo, reconstituted in MeOH/3% acetic acid in H_2O (30:70, 100 µl) and subjected to reversed phase and chiral column HPLC. Controls were performed using either denatured enzyme (boiled for 10 min) or in the absence of (±)-pinoresinols as substrate.

Lariciresinol reductase activity was assayed by monitoring the formation of (-)-[^{3}H]secoisolariciresinol. These assays were carried out exactly as described above, except that (±)-lariciresinols (5 mM in MeOH, 20 µl) were used as substrates, with (±)-secoisolariciresinols (20 µg) added as radiochemical carriers.

General Procedures for Enzyme Purification. Protein purification procedures were carried out at 4°C with chromatographic eluents monitored at 280 nm, unless otherwise stated. Protein concentrations were determined by the method of Bradford (Bradford, M.M., *Anal. Biochem.* 72:248-254 (1976)) using γ-globulin as standard. Polyacrylamide gel electrophoresis used gradient (4-15%, Bio-Rad) gels under denaturing and reducing conditions, these being performed in Laemmli's buffer system (Laemmli, U.K., *Nature* 227:680-685 (1970)). Proteins were visualized by silver staining (Morrissey, J.H., *Anal. Biochem.* 117:307-310 (1981)).

Preparation of crude extracts. F. intermedia stems (20 kg) were harvested, cut into 3-6 cm sections, and stored at -20°C until needed. Batches of stems (2 kg) were frozen in liquid nitrogen and pulverized in a Waring Blendor. The resulting powder was homogenized with potassium phosphate buffer (0.1 mM, pH 7.0, 4 L), containing 5 mM dithiothreitol. The homogenate was filtered through four layers of cheesecloth into a beaker containing 10% (w/v) polyvinylpolypyrolidone. The filtrate was centrifuged (12,000 x g, 15 min). The resulting supernatant was fractionated with $(NH_4)_2SO_4$, with proteins precipitating between 40 and 60% saturation recovered by centrifugation (10,000 x g, 1 h). The pellet was next reconstituted in a minimum amount of Tris-HCl buffer (20 mM, pH 8.0), containing 5 mM dithiothreitol (buffer A) and desalted using prepacked PD-10 columns (Sephadex G-25 medium) equilibrated with buffer A.

Affinity (Affi Blue Gel) Chromatography. The crude enzyme preparation (191 mg in buffer A, 5 nmol h^{-1} mg^{-1}) was applied to an Affi Blue Gel column (2.6 x 70 cm) equilibrated in buffer A. After washing the column with 200 ml of buffer A, pinoresinol/lariciresinol reductase was eluted with a linear NaCl gradient

(1.5-5 M in 300 ml) in buffer A at a flow rate of 1 ml min^{-1}. Active fractions were stored (-80°C) until needed.

Hydrophobic Interaction Chromatography (Phenyl Sepharose). After thawing, ten preparations resulting from the Affi Blue chromatography step (150 mg, 51 nmol h^{-1} mg^{-1}) were combined and applied to a Phenyl Sepharose column (1 x 10 cm) equilibrated in buffer A, containing 5 M NaCl. The column was washed with two bed volumes of the same buffer. Pinoresinol/lariciresinol reductase was eluted using a linear gradient of decreasing concentration of NaCl (5-0 M in 40 ml) in buffer A at a flow rate of 1 ml min^{-1}. Fractions catalyzing pinoresinol/lariciresinol reduction were combined and pooled.

Hydroxyapatite I Chromatography. Active protein (31 mg, 91 nmol h^{-1} mg^{-1}) from the phenyl sepharose purification step was applied to an hydroxyapatite column (1.6 x 70 cm) equilibrated in 10 mM potassium phosphate buffer, pH 7.0, containing 5 mM dithiothreitol (buffer B). Pinoresinol/lariciresinol reductase was eluted with a linear gradient of potassium phosphate buffer, pH 7.0 (0.01-0.4 M in 200 ml) at a flow rate of 1 ml min^{-1}. Active fractions were combined. The buffer was then exchanged with buffer A using PD-10 prepacked columns.

Affinity (2',5'-ADP Sepharose) Chromatography. The enzyme solution resulting from the hydroxyapatite purification step (6.5 mg, 463 nmol h^{-1} mg^{-1}) was next loaded on a 2',5'-ADP Sepharose (1 x 10 cm) column, previously equilibrated in buffer A containing 2.5 mM EDTA (buffer A') and then washed with 25 ml of buffer A'. Pinoresinol/lariciresinol reductase was eluted with a step gradient of NADP+ (0.3 mM in 10 ml) in buffer A' at a flow rate of 0.5 ml min^{-1}. [NAD+ (up to 3 mM) did not elute pinoresinol/lariciresinol reductase activity.] Because of the interference of the absorbance of the NADP+, it was not possible to directly monitor the eluent at 280 nm. Protein concentrations for each fraction were determined spectrophotometrically according to Bradford (Bradford, M.M., *Anal. Biochem.* 72:248-254 (1976)).

Hydroxyapatite II Chromatography. Fractions from the 2',5'-ADP Sepharose column that exhibited pinoresinol/lariciresinol reductase activity (0.85 mg, 1051 nmol h^{-1} mg^{-1}) were combined and directly applied to a second hydroxyapatite column (1 x 3 cm), equilibrated in buffer B, with the enzyme eluted with a linear gradient of potassium phosphate buffer, pH 7.0 (0.01-0.4 M in 45 ml) at a flow rate of 1 ml min^{-1}.

Affinity (Affi Yellow) Chromatography - Active fractions (160 µg, 7960 nmol h^{-1} mg^{-1}) from the second hydroxyapatite column purification step were next applied to a Reactive Yellow 3 Agarose column (1 x 3 cm), equilibrated in buffer A. Pinoresinol/lariciresinol reductase was eluted with a linear NaCl gradient (0-2.5 M in 100 ml) at a flow rate of 1 ml min^{-1}.

Fast Protein Liquid Chromatography (Superose 12 Chromatography) - Combined fractions from the Affi Yellow purification step having the highest activity (50 µg, 10,940 nmol h^{-1} mg^{-1}) were pooled and concentrated to 1 ml, using a Centricon 10 microconcentrator (Amicon, Inc.). The enzyme solution was then applied in portions of 200 µl to a fast protein liquid chromatography column (Superose 12, HR 10/30). Gel filtration was performed in a buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl and 5 mM dithiothreitol at a flow rate of 0.4 ml min^{-1}. Pinoresinol/lariciresinol reductase was eluted with 12.8 ml of the mobile phase. The active fractions which coincided with the UV profile (absorbance at 280 nm) were pooled (20 µg, 15,300 nmol h^{-1} mg^{-1}) and desalted (PD-10 prepacked columns).

The foregoing purification protocol resulted in a 3060-fold purification of (+)-pinoresinol/(+)-lariciresinol reductase. As for many of the enzymes involved in phenylpropanoid metabolism, the protein was in very low abundance, i.e. 20 kg *F. intermedia* stems yielded only ~20 µg of the purified (+)-pinoresinol/-(+)-lariciresinol reductase.

EXAMPLE 9

Characterization of Purified Pinoresinol/lariciresinol Reductases from *Forsythia Intermedia*

Isoelectric Focussing and pI Determination. In all stages of the purification protocol, (+)-pinoresinol/(+)-lariciresinol reductase activities coeluted. Given this observation, it was essential to unambiguously ascertain whether more than one form of the protein existed, *i.e.*, whether one form of the protein catalyzed the reduction of pinoresinol, and another form of the protein catalyzed the reduction of lariciresinol. To this end, the isoelectric point of pinoresinol/lariciresinol reductase was estimated by chromatofocussing on a MonoP HR 5/20 FPLC column.

Active fractions from the Superose 12 gel filtration column (Example 1) were pooled and the buffer exchanged with 25 mM Bis-Tris, pH 7.1, using prepacked PD-10 columns, equilibrated in the same buffer. The preparation so obtained was loaded on the chromatofocussing column and a pH gradient between 7.1 and 3.9 was

formed, using Polybuffer 74 as eluent at a flow rate of 0.5 ml min^{-1}. Aliquots (200 µl) of each fraction were assayed for pinoresinol/lariciresinol reductase activities. The remainder of the fractions was used to determine the pH gradient.

Molecular Weight Determination. Application of the MonoP HR 5/20 FPLC column preparation of pinoresinol/lariciresinol reductase to SDS-gradient gel electrophoresis (4-15% polyacrylamide) revealed the presence of two protein bands of similar apparent molecular weight, whose separation was achieved via anion-exchange chromatography on a MonoQ HR 5/5 FPLC matrix. Pooled fractions from the Sepharose 12 purification step (Example 1) were applied to a MonoQ HR 5/5 column (Pharmacia), equilibrated in buffer A. The column was washed with 10 ml of buffer A and pinoresinol/lariciresinol reductase activity eluted using a linear NaCl gradient (0-500 mM in 50 ml) in buffer A at a flow rate of 0.5 ml min^{-1}. Aliquots (30 µl) of the collected fractions were analyzed by SDS polyacrylamide gel electrophoresis, using a gradient (4-15% acrylamide) gel. Proteins were visualized by silver staining. Active fractions 34 through 37 (27,760 nmol h^{-1} mg^{-1}) and 38 through 41 (30,790 nmol h^{-1} mg^{-1}) were pooled separately and immediately used for characterization.

The two protein bands thus resolved under denaturing conditions had apparent molecular masses of ~36 and ~35 kDa, respectively. Each of the two reductase forms had a pI~5.7.

Native molecular weights of each reductase isoform were estimated via comparison of their elution behavior on Superose 12, Superose 6 and Superdex 75 gel filtration FPLC columns with the elution behavior of calibrated molecular weight standards. Gel filtration was carried out as set forth in Example 8. For each reductase, an apparent native molecular weight of 59,000 was calculated based on its elution volume, in contrast to that of ~36,000 and ~35,000 by SDS-polyacrylamide gel electrophoresis. While the discrepancy between molecular weights from gel filtration and SDS-PAGE remains unknown, it can tentatively be proposed that although the native protein likely exists as a dimer, it could also be a monomer of asymmetric shape, thereby altering its effective Stokes radius (Cantor, C.R., and Shimmel, P.R., *Biophysical Chemistry*, Part II, W.H. Freeman and Company, San Francisco, CA (1980); Stellwagen, E., *Methods in Enzymology* 182:317-328 (1990)); as reported for human thioredoxin reductase (Oblong, J.E., et al., *Biochemistry* 32:7271-7277 (1993)) and yeast metalloendopeptidase (Hrycyna, C.A., and Clarke, S., *Biochemistry* 32:11293-11301 (1993)).

pH and Temperature Optima. To determine the pH-optimum of pinoresinol/lariciresinol reductase, the enzyme preparation from the gel Superose 12 filtration step (Example 8) was assayed utilizing standard assay conditions (Example 8), except that the buffer was replaced with 50 mM Bis-Tris Propane buffer in the pH range of 6.3 to 9.4. The pH optimum was found to be pH 7.4.

The temperature optimum of pinoresinol/lariciresinol reductase was examined in the range between 4°C and 80°C under standard assay conditions (Example 8) utilizing the enzyme preparation from the gel filtration step (Example 8). At optimum pH, the temperature optimum for the reductase activity was established to be ~30°C.

Kinetic Parameters. Velocity studies were carried out to ascertain whether the two reductase isoforms catalyzed distinct reductions, i.e., that of the conversion of (+)-pinoresinol to (+)-lariciresinol, and (+)-lariciresinol to (-)-secoisolariciresinol, respectively, or whether either displayed a preference for (+)-pinoresinol or (+)-lariciresinol as substrates. The initial velocity studies were carried out individually utilizing the two isoforms of the enzyme, and individually employing both (+)-pinoresinol and (+)-lariciresinol as substrates. Initial velocity studies were performed in triplicate experiments, using 50 mM Bis-Tris Propane buffer, pH 7.4 containing 5 mM dithiothreitol, pure enzyme (after MonoQ anion-exchange chromatography), ten different substrate concentrations (between 8.8 and 160 µM) at a constant NADPH concentration (80 µM). Incubations were carried out at 30 °C for 10 min (within the linear kinetic range). Kinetic parameters were determined from Lineweaver-Burk plots.

Importantly, the kinetic parameters were essentially the same for both the 35 kDa and the 36 kDa forms of the enzyme (i.e., Km for pinoresinol: 27±1.5µm for the 35 kDa form of the enzyme, and 23±1.3µM for the 36 kDa form of the enzyme; Km for lariciresinol: 121±5.0µM for the 35 kDa form of the enzyme and 123±6.0µM for the 36 kDa form of the enzyme). In an analogous manner, apparent maximum velocities (expressed as µmol h^{-1} mg^{-1} of protein) were also essentially identical (i.e., Vmax for pinoresinol: 16.2±0.4 for the 35 kDa form of the enzyme and 17.3±0.5 for the 36 kDa form of the enzyme; for lariciresinol: 25.2±0.7 for the 35 kDa form of the enzyme and 29.9±0.7 for the 36 kDa form of the enzyme). Thus, all available evidence suggests that (+)-pinoresinol/(+)-lariciresinol reductase exists as two isoforms, with each capable of catalyzing the reduction of both substrates. How this reduction is carried out, i.e., whether both reductions are done in tandem, in either

quinone or furano ring form, awaits further study using a more abundant protein source.

Enzymatic Formation of (+)-[7'R-^{2}H]Lariciresinol. Since the two (+)-pinoresinol/(+)-lariciresinol reductase isoforms exhibited essentially identical catalytic characteristics, the Sepharose 12 enzyme preparation (Example 8), containing both isoforms, was used to examine the stereospecificity of the hydride transfer. The strategy adopted utilized selective deuterium labeling using NADP^{2}H as cofactor for the reduction of (+)-pinoresinol, with the enzymatic product, (+)-lariciresinol, being analyzed by ^{1}H NMR and mass spectroscopy. Thus, a solution of (±)-pinoresinols (5.2 mM in MeOH, 4 ml) was added to Tris-HCl buffer (20 mM, pH 8.0, containing 5 mM dithiothreitol, 22 ml) and stereospecifically deutero-labeled [4R-^{2}H]NADPH (20 mM in H_2O, 4 ml) prepared via the method of Anderson and Lin (Anderson, J.A., and Lin B.K., *Phytochemistry* **32**:811-812 (1993)), with the whole added to the enzyme preparation (20 ml). After incubation at 30°C for 1h with shaking, the assay mixture was extracted with EtOAc (2 x 50 ml). The EtOAc soluble fraction was combined, washed with saturated NaCl (50 ml), dried (Na_2SO_4), and evaporated to dryness in vacuo. The resulting extract was reconstituted in a minimum amount of EtOAc, applied to a silica gel column (0.5 x 7 cm), and eluted with EtOAc/hexanes (1:2). Fractions containing the enzymatic product were combined and evaporated to dryness.

The enzymatic product was established to be (+)-[7'R-^{2}H]lariciresinol, as evidenced by the disappearance of the 7'-proR proton at δ 2.51 ppm due to its replacement by deuterium and by its molecular ion at (m/z) 361 (M++1) corresponding to the presence of one deuterium atom at C-7. ^{1}H NMR (300 MHz) ($CDCl_3$): 2.39 (m,^{1}H, C8H), 2.71 (m,^{1}H, C8'H), 2.88 (δ,^{1}H, J7'S,8'=5.0 Hz, C7'HS), 3.73 (δδ,^{1}H, J8',9'b=7.0 Hz, J9'a,9'b=8.5 Hz, C9'Hß), 3.76 (δδ,^{1}H, J8,9S=6.5 Hz, J9R,9S=8.5 Hz, C9HS), 3.86 (s,^{3}H, OCH_3), 3.88 (s,^{3}H, OCH_3), 3.92 (δδ,^{1}H, J8,9R=6.0 Hz, J9R,9S=9.5 Hz, C9HR), 4.04 (δδ,^{1}H, J8',9'a=7.0 Hz, J9'a9'b=8.5 Hz, C9'Ha), 4.77 (δ,^{1}H, J7,8=6.6 Hz, C7H), 6.68 - 6.70 (m,^{2}H, ArH), 6.75 - 6.85 (m,4H, ArH); MS m/z (%) : 361 (M++1, 71.2), 360 (M+, 31.1), 237 (11.1), 153 (41.5), 152 (20.2), 151 (67.0), 138 (100), 137 (71.1).

Thus, hydride transfer from (+)-pinoresinol to (+)-lariciresinol had occurred in a manner whereby only the 7'-proR hydrogen position of (+)-lariciresinol was deuterated. An analogous result was observed for the conversion of (+)-lariciresinol

into (-)-secoisolariciresinol, thereby establishing that the overall hydride transfer was completely stereospecific.

EXAMPLE 10

Amino Acid Sequence Analysis of Purified Pinoresinol/Lariciresinol Reductase from *Forsythia intermedia*

Pinoresinol/Lariciresinol Reductase Amino Acid Sequencing. The (+)-pinoresinol/(+)-lariciresinol reductase N-terminal amino acid sequence was obtained from each of the purified proteins, and a mixture of both, using an Applied Biosystems protein sequencer with on-line HPLC detection. The N-terminal sequence was the same for both isoforms (SEQ ID No:36).

For trypsin digestion, 150 pmol of the enzyme purified from the Sepharose 12 column (Example 8) was suspended in 0.1 M Tris-HCl (50 µl, pH 8.5), with urea added to give a final concentration of 8 M in 77.5 µl. The mixture was incubated for 15 min at 50°C, then 100 mM iodoacetamide (2.5 µl) was added, with the whole kept at room temperature for 15 min. Trypsin (1 µg in 20 µl) was then added, with the mixture digested for 24 h at 37°C, after which TFA (4 µl) was added to stop the enzymatic reaction.

The resulting mixture was subjected to reversed phase HPLC analysis (C-8 column, Applied Biosytems), this being eluted with a linear gradient over 2 h from 0 to 100% acetonitrile (in 0.1% TFA) at a flow rate of 0.2 ml/min with detection at 280 nm. Fractions containing individual oligopeptide peaks were collected manually and directly submitted to amino acid sequencing. Four tryptic fragments were resolved in sufficient quantity to permit amino acid sequence determination. (SEQ ID Nos:37-40).

Cyanogen bromide digestion was performed by incubation of 150 pmol of the reductase purified from the Sepharose 12 column (Example 8) with 0.5 M cyanogen bromide in 70% formic acid for 40 h at 37°C, following which the cyanogen bromide and formic acid were removed by centrifugation under reduced pressure (SpeedVac). The resulting oligopeptide fragments were separated by HPLC and three were resolved in sufficient quantity to permit sequencing (SEQ ID Nos:41-43).

EXAMPLE 11

Cloning of Pinoresinol/Lariciresinol Reductase from *Forsythia intermedia*

Plant Materials. *Forsythia intermedia* plants were either obtained from Bailey's Nursery (var. Lynwood Gold, St., Paul, MN), and maintained in Washington State University greenhouse facilities, or were gifts from the local community.

Materials. All solvents and chemicals used were reagent or HPLC grade. UV RNA and DNA determinations at OD_{260} were obtained on a Lambda 6 UV/VIS spectrophotometer. A Temptronic II thermocycler (Thermolyne) was used for all PCR amplifications. Taq thermostable DNA polymerase was obtained from Promega, whereas restriction enzymes were from Gibco BRL (HaeIII), Boehringer Mannheim (Sau3a) and Promega (TaqI). pT7Blue T-vector and competent NovaBlue cells were purchased from Novagen and radiolabeled nucleotides ([α-^{32}P]dCTP and [γ-^{32}P]ATP) were from DuPont NEN.

Oligonucleotide primers for polymerase chain reaction (PCR) and sequencing were synthesized by Gibco BRL Life Technologies. GENECLEAN II® kits (BIO 101 Inc.) were used for purification of PCR fragments, with the gel-purified DNA concentrations determined by comparison to a low DNA mass ladder (Gibco BRL) in 1.5% agarose gels.

Forsythia RNA Isolation. Initial attempts to isolate functional *F. intermedia* RNA from fast-growing, green stem tissue were unsuccessful, due to difficulties encountered via facile oxidation by its plant phenolic constituents. This problem was, however, successfully overcome by utilization of an RNA isolation procedure, specifically designed for woody plant tissue, which uses low pH and reducing conditions in the extraction buffer to prevent oxidation (Dong, Z.D., and Dunstan, D.I., *Plant Cell Reports* 15: 516-521(1996)).

Forsythia intermedia stem cDNA Library Synthesis. Total RNA (~300 µg/g fresh weight) was obtained from young green stems of greenhouse-grown *Forsythia intermedia* plants (var. Lynwood Gold) (Dong, Z.D., and Dunstan, D.I., *Plant Cell Reports* 15:516-521 (1996)). A *Forsythia intermedia* stem cDNA library was constructed using 5 µg of purified poly A+ mRNA (Oligotex-dT™ Suspension, QIAGEN) with the ZAP-cDNA® synthesis kit, the Uni-ZAP™ XR vector and the Gigapack® II Gold packaging extract (Stratagene), with a titer of 1.2×10^6 PFU for the primary library. A portion (30 ml) of the amplified library (1.2×10^{10} PFU/ml; 158 ml total) was used to obtain pure cDNA library DNA for PCR (Sambrook, J. et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994); Ausubel, F.M. et al., *Current Protocols in Molecular Biology*, 2 volumes, Greene Publishing Associates and Wiley-Interscience, John Wiley & Sons, NY (1991)).

Pinoresinol/Lariciresinol Reductase DNA Probe Synthesis - The N-terminal and internal peptide amino acid sequences were used to construct the degenerate

oligonucleotide primers. Specifically, the primer PLRN5 (SEQ ID No:44) was based on the sequence of amino acids 7 to 13 of the N-terminal peptide (SEQ ID No:36). The primer PLR14R (SEQ ID No:45) was based on the sequence of amino acids 2 to 8 of the internal peptide sequence set forth in (SEQ ID No:37). The primer PLR15R (SEQ ID No:46) was based on the sequence of amino acids 9 to 15 of the internal peptide sequence set forth in (SEQ ID No:37). The sequence of amino acids 9 to 15 of the internal peptide sequence set forth in SEQ ID No:37, upon which the sequence of primer PLR15R (SEQ ID No:46) was based, also corresponded to the sequence of amino acids 4 to 10 of the cyanogen bromide-generated, internal fragment set forth in SEQ ID No:41.

Purified *F. intermedia* cDNA library DNA (5 ng) was used as the template in 100 µl PCR reactions (10 mM Tris-HCl [pH 9.0], 50 mM KCl, 0.1% Triton X-100, 2.5 mM $MgCl_2$, 0.2 mM each dNTP and 2.5 units Taq DNA polymerase) with primer PLRN5 (SEQ ID No:44) (100 pmol) and either primer PLRI5R (SEQ ID No:46) (20 pmol) or primer PLRI4R (SEQ ID No:45) (20 pmol). PCR amplification was carried out in a thermocycler as follows: 35 cycles of 1 min at 94°C, 2 min at 50°C and 3 min at 72°C; with 5 min at 72°C and an indefinite hold at 4°C after the final cycle. Single-primer, template-only and primer-only reactions were performed as controls. PCR products were resolved in 1.5% agarose gels. The combination of primers PLRN5 (SEQ ID No:44) and PLRI4R (SEQ ID No:45) yielded a single band of 380-bp corresponding to bases 22 to 393 of SEQ ID No:47. The combination of primers PLRN5 (SEQ ID No:44) and PLRI5R (SEQ ID No:46) yielded a single band of 400-bp corresponding to bases 22 to 423 of SEQ ID No:47.

To determine the nucleotide sequence of the two amplified bands, five, 100 µl PCR reactions were performed as above with each of the following combinations of template and primers: 380 bp amplified product plus primers PLRN5 (SEQ ID No:44) and PLRI4R (SEQ ID No:45); 400 bp amplified product plus primers PLRN5 (SEQ ID No:44) and PLRI5R (SEQ ID No:46). The 5 reactions from each combination of primers and template were concentrated (Microcon 30, Amicon Inc.) and washed with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA; 2 x 200 µl), with the PCR products subsequently recovered in TE buffer (2 x 50 µl). These were resolved in preparative 1.5% agarose gels. Each gel-purified PCR product (~0.2 pmol) was then ligated into the pT7Blue T-vector and transformed into competent NovaBlue cells, according to Novagen's instructions. Insert sizes were determined using the rapid boiling lysis and PCR technique (utilizing R20mer

(SEQ ID No:74) and U19mer (SEQ ID No:75) primers according to the manufacturer's (Novagen's) instructions.

Restriction analysis was performed to determine whether all inserts for each combination of primers and template were the same. Restriction analysis was carried out as follows: each of the inserts was amplified by PCR utilizing the R20 (SEQ ID No:74) and U19 (SEQ ID No:75) primers. To 20 µl each of a 100 µl PCR reaction were added 4 units HaeIII, 1.5 units Sau3a or 5 units TaqI restriction enzyme. Restriction digestions were allowed to proceed for 60 min at 37°C for HaeIII and Sau3A and at 65°C for TaqI reactions. Restriction products were resolved in 1.5% agarose gels giving one restriction group for all inserts tested.

Five of the resulting, recombinant plasmids were selected for DNA sequencing. The inserts from three of the recombinant plasmids (called pT7PLR1-pT7PLR3) were generated by a combination of primers PLRN5 (SEQ ID No:44) and PLRI5R (SEQ ID No:46) with the 400 bp PCR product as substrate. The inserts from the remaining two recombinant plasmids (called pT7PLR4 and pT7PLR5) were generated from a combination of primers PLRN5 (SEQ ID No:44) and PLRI4R (SEQ ID No:45) and the 380 bp PCR product as substrate. All of the five, sequenced PCR products contained the same open reading frame.

The (+)-pinoresinol/(+)-lariciresinol reductase probe was constructed as follows: five, 100 µl PCR reactions were performed as described above with 10 ng pT7PLR3 DNA with primers PLRN5 (SEQ ID No:44) and PLRI5R (SEQ ID No:46). Gel-purified pT7PLR3 cDNA insert (50 ng) was used with Pharmacia's T7QuickPrime® kit and [α-^{32}P]dCTP, according to kit instructions, to produce a radiolabeled probe (in 0.1 ml), which was purified over BioSpin 6 columns (Bio-Rad) and added to carrier DNA (0.9 ml of 0.5 mg/ml sheared salmon sperm DNA obtained from Sigma).

Library Screening. 600,000 PFU of *F. intermedia* amplified cDNA library were plated for primary screening, according to Stratagene's instructions. Plaques were blotted onto Magna Nylon membrane circles (Micron Separations Inc.), which were then allowed to air dry. The membranes were placed between two layers of Whatman® 3MM Chr paper. cDNA library phage DNA was fixed to the membranes and denatured in one step by autoclaving for 2 min at 100°C with fast exhaust. The membranes were washed for 30 min at 37°C in 6X standard saline citrate (SSC) and 0.1% SDS and prehybridized for 5 h with gentle shaking at 57-58°C in preheated 6X

SSC, 0.5% SDS and 5X Denhardt's reagent (hybridization solution, 300 ml) in a crystallization dish (190x75 mm).

The [^{32}P]radiolabeled probe was denatured (boiling, 10 min), quickly cooled (ice, 15 min) and added to a preheated fresh hybridization solution (60 ml, 58°C) in a crystallization dish (150x75 mm). The prehybridized membranes were next added to this dish, which was then covered with plastic wrap. Hybridization was performed for 18 h at 57-58°C with gentle shaking. The membranes were washed in 4X SSC and 0.5% SDS for 5 min at room temperature, transferred to 2X SSC and 0.5% SDS (at room temperature) and incubated at 57-58°C for 20 min with gentle shaking, wrapped with plastic wrap to prevent drying and finally exposed to Kodak X-OMAT AR film for 24 h at -80°C with intensifying screens.

This screening procedure resulted in more than 350 positive plaques, with twenty (of different signal intensities) being subjected to two additional rounds of screening. After final purification, six of the twenty cDNAs were subcloned by *in vivo* excision into pBluescript. These six cDNAs were called plr-Fi1 to plr-Fi6 (SEQ ID Nos:47, 49, 51, 53, 55, 57).

In vivo Excision and Sequencing of plr-Fi1-plr-Fi6 Phagemids. The six purified cDNA clones were rescued from the phage following Stratagene's *in vivo* excision protocol. Both strands of the six different cDNAs (plr-Fi1 to plr-Fi6) that coded for (+)-pinoresinol/ (+)-lariciresinol reductase were completely sequenced using overlapping sequencing primers.

Purification of DNA for sequencing employed a QIAwell Plus plasmid purification system (QIAGEN) followed by PEG precipitation (Sambrook, J., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)), with DNA sequences determined using an Applied Biosystems Model 373A automated sequencer. DNA and amino acid sequence analyses were performed using the Unix-based GCG Wisconsin Package (Program Manual for the Wisconsin Package, Version 8, September 1994, *Genetics Computer Group*, 575 Science Drive, Madison, Wisconsin, USA 53711; Rice, P., *Program Manual for the EGCG Package*, Peter Rice, The Sanger Centre, Hinxton Hall, Cambridge, CB10 1Rq, England (1996)) and the ExPASy World Wide Web molecular biology server (Geneva University Hospital and University of Geneva, Geneva, Switzerland).

All six cDNAs had the same coding but different 5'-untranslated regions. On the other hand, analysis of the 3'-untranslated region of each of the six cDNAs

established that all were truncated versions of the longest cDNA's 3'-region. Preliminary RNA gel blot analysis with total RNA from greenhouse-grown plant stem tips confirmed a single transcript with a length of approximately 1.2 kb.

RNA gel blot analysis. For RNA gel blot analysis, total RNA (30 µg per lane) from *F. intermedia* stem tips was separated by size by denaturing agarose gel electrophoresis. The RNA was transferred to charged nylon membranes (GeneScreen Plus®, Dupont NEN), cross-linked to the membrane (Stratalinker from Stratagene), prehybridized, hybridized with the same probe used to screen the cDNA library during cDNA cloning and washed according to the manufacturer's instructions for aqueous hybridization conditions. The membrane was then exposed to Kodak X-OMAT film for 48 hr at -80°C with intensifying screens.

EXAMPLE 12

Expression of (+)-Pinoresinol/(+)-Lariciresinol Reductase cDNA plr-Fi1 in *E.coli*

Expression in Escherichia coli. In order to confirm that the putative (+)-pinoresinol/(+)-lariciresinol reductase cDNAs encoded functional (+)-pinoresinol/(+)-lariciresinol reductase, the cDNAs putatively encoding (+)-pinoresinol/(+)-lariciresinol reductase were heterologously expressed in *E. coli.* Heterologous expression was also necessary in order to obtain sufficient protein to enable the systematic study of the precise biochemical mechanism of (+)-pinoresinol/(+)-lariciresinol reductase at a future date.

Examination of the six putative (+)-pinoresinol/(+)-lariciresinol reductase clones revealed that one, plr-Fi1 (SEQ ID No:47), was in frame with the α-complementation particle of β-galactosidase in pBluescript. This was fortuitous, since it potentially provided a facile means to express the fully functional fusion protein, and hence to provide proof that the cloned sequence was correct.

Purified plasmid DNA from plr-Fi1 (SEQ ID No:47) was transformed into NovaBlue cells according to Novagen's instructions. Transformed cells (5 ml cultures) were grown at 37°C with shaking (225 rpm) to mid log phase (OD_{600}=0.5) in LB medium (Sambrook, J., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)) supplemented with 12.5 µg ml^{-1} tetracycline and 50 µg ml^{-1} ampicillin. IPTG (isopropyl β-D-thioglucopyranoside) was then added to a final concentration of 10 mM, and the cells were allowed to grow for 2 h. Cells were collected by centrifugation and resuspended in 500 µl (per 5 ml culture tube) buffer (20 mM Tris-HCl, pH 8.0, 5 mM dithiothreitol). Lysozyme (5 µl of 0.1 mg ml^{-1}, Research

Organics, Inc.) was next added and following incubation for 10 min, the cells were lysed by sonication (3 x 15 s). After centrifugation at 14,000 x g at 4°C for 10 min, the supernatant was removed and assayed for (+)-pinoresinol/(+)-lariciresinol reductase activity (210 µl supernatant per assay) as described in Example 8.

Catalytic activity was established by incubating cell-free extracts for 2 h at 30°C with (±)-pinoresinols (0.4 mM) and [4R-^{3}H]NADPH (0.8 mM) under standard conditions. Following incubation, unlabeled (±)-lariciresinols and (±)-secoisolariciresinols were added as radiochemical carriers, with each lignan isolated by reversed-phase HPLC. Controls included assays of a pinoresinol/lariciresinol reductase cDNA which contains an out-of-frame cDNA insert, with all assay components, as well as plr-Fi1 (SEQ ID No:47) and an out-of-frame pinoresinol/-lariciresinol reductase cDNA with no substrate except [4R-^{3}H]NADPH. Separation of products and chiral identification were performed by HPLC as previously described (Chu, A., et al., *J. Biol. Chem.* 268:27026-27033 (1993)).

Subsequent chiral HPLC analysis revealed that both (+)-lariciresinol and (-)-secoisolariciresinol, but not the corresponding antipodes, were radiolabeled (total activity: 54 nmol h^{-1} mg^{-1}). By contrast, no catalytic activity was detected either in the absence of (±)-pinoresinols, or when control cells were used which contained a plasmid in which the cDNA insert was not in-frame with the ß-galactosidase gene. Thus, the heterologously expressed (+)-pinoresinol/(+)-lariciresinol reductase and the plant protein function in precisely the same enantiospecific manner.

EXAMPLE 13

Sequence and Homology Analysis of the cDNA Insert of Clone plr-Fi1 (SEQ ID No:47) Encoding (+)-pinoresinol/(+)-lariciresinol reductase

Sequence Analysis. The full length sequence of the cloned (+)-pinoresinol/(+)-lariciresinol reductase plr-Fi1(SEQ ID No:47) contained all of the peptide sequences determined by Edman degradation of digest fragments.

The single ORF predicts a polypeptide of 312 amino acids (SEQ ID No:48) with a calculated molecular mass of 34.9 kDa, in close agreement with the value (~35 or ~36 kDa) estimated previously by SDS-PAGE for the two isoforms of (+)-pinoresinol/(+)-lariciresinol reductase. An equal number of acidic and basic residues are also present, with a theoretical isoelectric point (pI) of 7.09, in contrast to that experimentally obtained by chromatofocussing (pI ~5.7).

The amino acid composition reveals seven methionine residues. Interestingly, the N-terminus of the plant-purified enzyme lacks the initial

methionine, this being the most common post-translational protein modification known. Consequently, the first methionine in the cDNA can be considered to be the site of translational initiation. The sequence analysis also reveals a possible N-glycosylation site at residue 215 (although no secretory targeting signal is present), and seven possible protein phosphorylation sites at residues 50 and 228 (protein kinase C-type), residues 228, 250, 302 and 303 (casein kinase II-type) and residue 301 (tyrosine kinase type).

Regions of the pinoresinol/lariciresinol polypeptide chain (SEQ ID NO:48) were also identified that contained conserved sequences associated with NADPH binding (Jörnvall, H., in *Dehydrogenases Requiring Nicotinamide Coenzymes* (Jeffery, J., ed) pp. 126-148, Birkhäuser Verlag, Basel (1980); Branden, C., and Tooze, J., *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York and London (1991); Wierenga, R.K. et al., *J. Mol. Biol.* 187:101-108 (1986)). There is a limited number of invariant amino acids in the sequences of different reductases which are viewed as indicative of NADPH binding sites. These include three conserved glycine residues with the sequence G-X-G-X-X-G (SEQ ID No:76), where X is any residue, and six conserved hydrophobic residues. The glycine-rich region is considered to play a central role in positioning the NADPH in its correct conformation. In this regard, a comparison of the N-terminal region of (+)-pinoresinol/(+)-lariciresinol reductase with that of the conserved, NADPH-binding regions of *Drosophila melanogaster* alcohol dehydrogenase (Branden, C., and Tooze, J., *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York and London (1991)), *Pinus taeda* cinnamyl alcohol dehydrogenase (MacKay J.J. et al., *Mol. Gen. Genet.* 247:537-545 (1995)), dogfish muscle lactate dehydrogenase (Branden, C., and Tooze, J., *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York and London (1991)) and human erythrocyte glutathione reductase (Branden, C., and Tooze, J., *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York and London (1991)), revealed some interesting parallels. The invariant glycine residues are aligned in every case, as are four of the six hydrophobic residues required for the correct packaging in the formation of the domain. Hence, the NADPH-binding site of (+)-pinoresinol/(+)-lariciresinol reductase isoforms is localized close to the N-terminus.

Homology Analysis: Comparison to Isoflavone Reductase. A BLAST search (Altschul, S.F, et al., *J. Mol. Biol.* 215:403-410 (1990)) was conducted with the

translated amino acid sequence of (+)-pinoresinol/(+)-lariciresinol reductase (SEQ ID No:48) against the non-redundant peptide database at the National Center for Biotechnology Information. Significant homology was noted for (+)-pinoresinol/(+)-lariciresinol reductase with various isoflavone reductases from the legumes, *Cicer arietinum* (Tiemann, K., et al., *Eur. J. Biochem.* 200:751-757 (1991)) (63.5% similarity, 44.4% identity), *Medicago sativa* (Paiva, N.L., et al., *Plant Mol. Biol.* 17:653-667 (1991)) (62.6% similarity, 42.0% identity) and *Pisum sativum* (Paiva, N.L., et al., *Arch. Biochem. Biophys.* 312:501-510 (1994)) (61.6% similarity, 41.3% identity). This observation is of considerable interest since isoflavonoids are formed via a related branch of phenylpropanoid-acetate pathway metabolism. Specifically, isoflavone reductases catalyze the reduction of α,β-unsaturated ketones during isoflavonoid formation. For example, the *Medicago sativa L.* isoflavone reductase catalyzes the stereospecific conversion of 2'-hydroxy-formononetin to (3R)-vestitone in the biosynthesis of the phytoalexin, (-)-medicarpin (Paiva, N.L. et al., *Plant Mol. Biol.* 17:653-667 (1991)). This sequence similarity may be significant given that both lignans and isoflavonoids are offshoots of general phenylpropanoid metabolism, with comparable plant defense functions and pharmacological roles, e.g., as "phytoestrogens". Consequently, since both reductases catalyze very similar reactions, it is tempting to speculate that the isoflavone reductases may have evolved from (+)-pinoresinol/(+)-lariciresinol reductase. This is considered likely since the lignans are present in the pteridophytes, hornworts, gymnosperms and angiosperms; hence their pathways apparently evolved prior to the isoflavonoids (Gang et al., In *Phytochemicals for Pest Control*, Hedin et al., eds, ACS Symposium Series, Washington D.C., 658:58-59 (1997)).

Comparable homology was also observed with putative isoflavone reductase "homologs" from *Arabidopsis thaliana* (Babiychuk, E., et al., Direct Submission (25-MAY-1995) to the EMBL/GenBank/DDBJ databases (1995)) (65.9% similarity, 50.8% identity), *Nicotiana tabacum* (Hibi, N., et al., *Plant Cell* 6:723-735 (1994)) (64.6% similarity, 47.2% identity), *Solanum tuberosum* (van Eldik, G.J., et al., (1995) Direct submission (06-OCT-1995) to the EMBL/GenBank/DDBJ databases) (65.5% similarity, 47.7% identity) *Zea mays* (Petrucco, S., et al., *Plant Cell* 8:69-80 (1996)) (61.6% similarity, 44.9% identity) and especially *Lupinus albus* (Attuci, S., et al., Personal communication and direction submission (06/6/96) to the EMBL/Genbank/DDBJ databases (1996)) (85.9% similarity, 66.2% identity).

By contrast, homology with other NADPH-dependent reductases was significantly lower: for example, dihydroflavonol reductases from *Petunia hybrida* (Beld, M. et al., *Plant Mol. Biol.* 13:491-502 (1989)) (43.2% similarity, 21.5% identity) and *Hordeum vulgare* (Kristiansen, K.N., and Rohde, W., *Mol. Gen. Genet.* 230:49-59 (1991)) (46.2% similarity, 21.1% identity), chalcone reductase from *Medicago sativa* (Ballance, G.M. and Dixon, R.A., *Plant Physiol.* 107:1027-1028 (1995)) (39.5% similarity, 15.8% identity), chalcone reductase "homolog" from *Sesbania rostrata* (Goormachtig, S., et al., (1995) Direct Submission (13-MAR-1995) to the EMBL/GenBank/DDBJ databases) (47.6% similarity, 24.1% identity), cholesterol dehydrogenase from *Nocardia sp.* (Horinouchi, S., et al., *Appl. Environ. Microbiol.* 57:1386-1393 (1991)) (46.6% similarity, 21.0% identity) and 3-β-hydroxy-5-ene steroid dehydrogenase from *Rattus norvegicus* (Zhao, H.-F., et al., *Journal Endocrinology* 127:3237-3239 (1990)) (43.5% similarity, 20.6% identity).

Thus, sequence analysis establishes significant homology between (+)-pinoresinol/(+)-lariciresinol reductase, isoflavone reductases and putative isoflavone reductase "homologs" which do not possess isoflavone reductase activity.

EXAMPLE 14

cDNA Cloning of *Thuja plicata* (-)-Pinoresinol/(-)-Lariciresinol Reductases

Plant Materials. Western red cedar plants (*Thuja plicata*) were maintained in Washington State University greenhouse facilities.

Materials. All solvents and chemicals used were reagent or HPLC grade. *Taq* thermostable DNA polymerase and restriction enzymes (SacI and XbaI) were obtained from Promega. pT7Blue T-vector and competent NovaBlue cells were purchased from Novagen and radiolabeled nucleotide ([α-^{32}P]dCTP) was purchased from DuPont NEN.

Oligonucleotide primers for polymerase chain reaction (PCR) and sequencing were synthesized by Gibco BRL Life Technologies. GENECLEAN II® kits (BIO 101 Inc.) were used for purification of PCR fragments, with the gel-purified DNA concentrations determined by comparison to a low DNA mass ladder (Gibco BRL) in 1.3% agarose gels.

Instrumentation. UV (including RNA and DNA determinations at OD_{260}) spectra were recorded on a Lambda 6 UV/VIS spectrophotometer. A Temptronic II thermocycler (Thermolyne) was used for all PCR amplifications. Purification of plasmid DNA for sequencing employed a QIAwell Plus plasmid purification system (Qiagen) followed by PEG precipitation (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)) or Wizard® Plus SV Minipreps DNA Purification System (Promega), with DNA sequences determined using an Applied Biosystems Model 373A automated sequencer.

Thuja plicata cDNA Library Synthesis. Total RNA (6.7 μg/g fresh weight) was obtained from young green leaves (including stems) of greenhouse-grown western red cedar plants (*Thuja plicata*) according to the method of Lewinsohn et al (Lewinsohn, E., et al., *Plant Mol. Biol. Rep.* 12:20-25 (1994)). A *T.plicata* cDNA library was constructed using 3 μg of purified poly(A)+ mRNA (Oligotex-dT™ Suspension, Qiagen) with the ZAP-cDNA® synthesis kit, the Uni ZAP™ XR vector, and the Gigapack® II Gold packaging extract (Stratagene), with a titer of 1.2 X 10^5 pfu for the primary library. The amplified library (7.1 X 10^8 pfu /ml; 28 ml total) was used for screening (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)).

T. plicata (-)-Pinoresinol/(-)-Lariciresinol Reductase cDNA Synthesis. *T. plicata* (-)-pinoresinol/(-)-lariciresinol reductase cDNA was obtained from mRNA by a reverse transcription-polymerase chain reaction (RT-PCR) strategy (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)). First-strand cDNA was synthesized from the purified mRNA previously used for the synthesis of the *T. plicata* cDNA library, described above. Purified mRNA (150 ng) was mixed with linker-primer (1.4 μg) from ZAP-cDNA® synthesis kit (Stratagene), heated to 70°C for 10 min, and quickly chilled on ice. The mixture of denatured mRNA template and linker-primer was then mixed with First Strand Buffer (Life Technologies), 10 mM DTT, 0.5 mM each dNTP, and 200 units of Super Script™II (Life Technologies) in a final volume of 20 μl. The reaction was carried out at 42°C for 50 min and then stopped by heating (70°C, 15 min). *E. coli* RNase H (1.5 units, 1 μl) was added to the solution and incubated at 37°C for 20 min.

The first-strand reaction (2 μl) was next used as the template in 100-μl PCR reactions (10 mM Tris-HCl, pH 9.0, 50 mM KCl, 0.1 % Triton X-100, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, and 5 units of Taq DNA polymerase) with primer CR6-NT (5'GCACATAAGAGTATGGATAAG3')(SEQ ID No:60) (10 pmol) and primer XhoI-Poly(dT) (5'GTCTCGAGTTTTTTTTTTTTTTTTTT3')(SEQ ID No:59) (10 pmol). PCR amplification was carried out in a thermocycler as described in

(Dinkova-Kostova, A.T., et al., *J. Biol. Chem.* 271:29473-29482 (1996)) except for the annealing temperature at 52°C. PCR products were resolved in 1.3 % agarose gels, where at least two bands possessing the expected length (about 1,200-bp) were observed. The bands were extracted from the gel. The gel-purified PCR products (56 ng) were then ligated into the pT7Blue T-vector (50 ng) and transformed into competent NovaBlue cells, according to Novagen's instructions.

The size and orientation of the inserted cDNAs were determined using the rapid boiling lysis and PCR technique, following the manufacturer's (Novagen's) instructions, with the following primer combinations: R20-mer(SEQ ID No:74) with U19-mer (SEQ ID No:75); R20-mer (SEQ ID No:74) with CR6-NT (SEQ ID No:60); U19-mer (SEQ ID No:75) with CR6-NT (SEQ ID No:60). The CR6-NT primer end of the inserted DNAs was located next to the U19-mer primer site of the T-vector. The T-vectors containing the inserted cDNAs were purified with Wizard® Plus SV Minipreps DNA Purification System. Five inserted cDNAs were completely sequenced using overlapping sequencing primers and were shown to be identical except that polyadenylation sites were different. Therefore, the longest cDNA, designated plr-Tp1, (SEQ ID No:61) was used for detection of enzyme activity using the pBluescript expression system.

Sequence Analysis - DNA and amino acid sequence analyses were performed using the Unix-based GCG Wisconsin Package (Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 (1996); Rice, P., Program Manual for the EGCG Package, Peter Rice, The Sanger Centre, Hinxton Hall, Cambridge, CB10 1Rq, England) and the ExPASy World Wide Web molecular biology server (Geneva University Hospital and University of Geneva, Geneva, Switzerland).

EXAMPLE 15

cDNA Cloning and Expression of *Thuja plicata* (+)-Pinoresinol/ (+)-Lariciresinol Reductase

T. plicata (+)-Pinoresinol/(+)-Lariciresinol Reductase cDNA cloning. After plr-Tp1 was cloned and sequenced, the full-length clone was used to screen the *T. plicata* cDNA library as described in Example 11, except that the entire plr-Tp1 cDNA insert was used as a probe. Several positive clones were sequenced, revealing one new, unique cDNA which was called plr-Tp2. This cDNA encodes a reductase with high sequence similarity to plr-Tp1 (~81% similarity at the amino acid level), but with substrate specificity properties identical to the original *Forsythia intermedia* reductase, as described below.

Enzyme Assays. Pinoresinol and lariciresinol reductase activities were assayed by monitoring the formation of [^{3}H]lariciresinol and [^{3}H]secoisolariciresinol as set forth in Example 8, with the following modifications. Briefly, each assay for pinoresinol reductase activity consisted of (±)-pinoresinols (5 mM in MeOH, 20 µl) and the enzyme preparation (i.e., total protein extract from *E. coli*, 210 µl). The enzymatic reaction was initiated by addition of [4R-^{3}H]NADPH (10 mM, 6.79 kBq/mmol in distilled H_2O, 20 µl). After 3 hour incubation at 30°C with shaking, the assay mixture was extracted with EtOAc (500 µl) containing (±)-lariciresinols (20 µg) and (±)-secoisolariciresinols (20 µg) as radiochemical carriers. After centrifugation (13,800 x g, 5 min), the EtOAc solubles were removed and the extraction procedure was repeated. For each assay, the EtOAc solubles were combined with an aliquot (100 µl) removed for determination of its radioactivity using liquid scintillation counting. The remainder of the combined EtOAc solubles was evaporated to dryness *in vacuo*, reconstituted in MeOH/H_2O (30:70, 100 µl) and subjected to reversed phase and chiral column HPLC.

Lariciresinol reductase activity was assayed by monitoring the formation of (+)-[^{3}H]secoisolariciresinol. These assays were carried out exactly as described above, except that (±)-lariciresinols (5 mM in MeOH, 20 µl) were used as substrates, with (±)-secoisolariciresinols (20 µg) added as radiochemical carriers.

Expression of plr-Tp1 in E. coli - In order for the open reading frame (ORF) of plr-Tp1 to be in frame with the β-galactosidase gene α-complementation particle in pBluescript SK(-), plr-Tp1 was excised out of pT7Blue T-vector with SacI and XbaI, gel-purified, and then ligated into the expression vector digested with these same enzymes. This plasmid, pPCR-Tp1, was transformed into NovaBlue cells according to Novagen's instructions. The transformed cells (5-ml cultures) were grown at 37°C in LB medium (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, 3 volumes, 3rd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1994)) supplemented with 50 µg ml^{-1} carbenicillin with shaking (225 rpm) to mid log phase (A_{600} = 0.5-0.7). The cells were next collected by centrifugation (1000 x g, 10 min) and resuspended in fresh LB medium supplemented with 10 mM IPTG (isopropyl β-D-thioglucopyranoside) and 50 µg ml^{-1} carbenicillin to an absorbance of 0.6 (at 600 nm). The cells, allowed to grow overnight, were collected by centrifugation and resuspended in 500-700 µl of (per

5 ml culture tube) of buffer (50 mM Tris-HCl, pH 7.5, 2 mM EDTA, and 5 mM DTT). Next, the cells were lysed by sonication (5 x 45 s) and after centrifugation (17500 x g, 4°C, 10 min) the supernatant was removed and assayed for (-)-pinoresinol/(-)-lariciresinol reductase activity as described above. Controls included assays of pBluescript (SK(-)) without insert DNA (as negative control) or with pPLR-Fi1 (cDNA of authentic *F. intermedia* (+)-pinoresinol/ (+)-lariciresinol reductase in frame) as stereospecific control, as well as pPLR-Tp1 with no substrate except (4R)-^{3}HNADPH.

The results showed that both (-)-lariciresinol and (+)-secoisolariciresinol were radiolabeled and that no incorporation of radioactivity was found in (-)-secoisolariciresinol. However, accumulation of radiolabel into (+)-lariciresinol was also observed, although at a much slower rate than that observed for (-)-lariciresinol. These results indicate that plr-Tp1 can use both (-)-pinoresinol and (+)-pinoresinol as substrates, with the former being converted via (-)-lariciresinol completely to (+)-secoisolariciresinol, and the latter being converted much more slowly to (+)-lariciresinol, but not further to (-)-secoisolariciresinol.

Expression of plr-Tp2 *in E. coli.* The plr-Tp2 cDNA was found to be in frame with the β-galactosidase gene α-complementation particle in pBluescript SK(-). When evaluated for activity and substrate specificity, as described above, plr-Tp2 was found to possess the same substrate specificity and product formation as the original *Forsythia intermedia* reductase (Dinkova-Kostova, A.T., et al., *J. Biol. Chem.* 271:29473-29482 (1996)) except that a small amount of (-)-lariciresinol was also detected. This is interesting, because plr-Tp2 has a higher sequence similarity to plr-Tp1 than it does to the *Forsythia* reductase.

All the above observations were confirmed using deuterolabeled substrates (±)-[9,9'-$^2H_2$, OC^2H_3]pinoresinols with isolation of the corresponding lignans; each was then subjected to chiral column chromatography and HPLC-mass spectral analysis to confirm these findings.

EXAMPLE 16

Cloning of Additional Pinoresinol/Lariciresinol Reductases from *Thuja plicata* and *Tsuga heterophylla*

Two additional pinoresinol/lariciresinol reductases were cloned from a *Thuja plicata* young stem cDNA library as described in Example 15 for the cloning of plr-Tp2. The two additional pinoresinol/lariciresinol reductases were designated plr-Tp3 (SEQ ID No:65) and plr-Tp4 (SEQ ID No:67).

Two additional pinoresinol/lariciresinol reductases were cloned from a *Tsuga heterophylla* young stem cDNA library as described in Example 15 for the cloning of plr-Tp2. The two additional pinoresinol/lariciresinol reductases from *Tsuga heterophylla* were designated plr-Tp3 (SEQ ID No:69) and plr-Tp4 (SEQ ID No;71).

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Lewis, Norman G
Davin, Laurence B
Dinkova-Kostova, Albena T
Fujita, Masayuki
Gang, David R
Sarkanen, Simo

(ii) TITLE OF INVENTION: Recombinant Pinoresinol/Lariciresinol Reductases,Recombinant Dirigent Proteins and Methods of Use

(iii) NUMBER OF SEQUENCES: 76

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Christensen, O'Connor, Johnson & Kindness
(B) STREET: 1420 Fifth Avenue, Suite 2800
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: WA 98101-2347

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Shelton, Dennis K
(B) REGISTRATION NUMBER: 26,997
(C) REFERENCE/DOCKET NUMBER: WSUR111351

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 206 682 8100
(B) TELEFAX: 206 224 0779

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Forsythia intermedia dirigent protein N-terminal sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Lys Pro Arg Pro Xaa Arg Xaa Xaa Lys Glu Leu Val Phe Tyr Phe Xaa
1               5                   10                  15

Asp Ile Leu Phe Lys Gly Xaa Asn Tyr Asn Xaa Ala
            20                  25
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Thr Ala Met Ala Val Pro Phe Asn Tyr Gly Asp Leu Val Val Phe Asp
1               5                   10                  15

Asp Pro Ile Thr Leu Asp Asn Asn
            20
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Tyr Val Gly Thr Leu Asn Phe Ala Gly Ala Asp Pro Leu Leu Xaa Lys
1               5                   10                  15
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Asp Ile Ser Val Ile Gly Gly Thr Gly Asp Phe Phe Met Ala Arg
1               5                   10                  15
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Gly Val Ala Thr Leu Met Thr Asp Ala Phe Glu Gly Asp Xaa Tyr
1               5                   10                  15
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Ala Gln Gly Met Tyr Phe Tyr Asp Gln Lys
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: Forsythia intermedia dirigent protein internal tryptic fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Tyr Asn Ala Trp Leu
1               5
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: "PCR primer PSINT1"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
AARGARYTNG TNTTYTAYTT Y                                              21
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: "PCR primer PSI1R"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
TARTTRAANG GNACNGCCAT                                              20
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION:"PCR primer PSI2R"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
GTNATNGGRT CRTCRAANAC                                              20
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION:"PCR primer PSI7R"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
CCATRAARAA RTCNCCNGT                                               19
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 901 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Forsythia intermedia clone psd-fil

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 26..583

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
ATTTCGGCAC GAGATTAAAC CAAAC ATG GTT TCT AAA ACA CAA ATT GTA GCT      52
                            Met Val Ser Lys Thr Gln Ile Val Ala
                              1               5

CTT TTC CTT TGC TTC CTC ACT TCC ACC TCT TCC GCC ACC TAC GGC CGC     100
Leu Phe Leu Cys Phe Leu Thr Ser Thr Ser Ser Ala Thr Tyr Gly Arg
 10                  15                  20                  25

AAG CCA CGC CCT CGC CGG CCC TGC AAA GAA TTG GTG TTC TAT TTC CAC     148
Lys Pro Arg Pro Arg Arg Pro Cys Lys Glu Leu Val Phe Tyr Phe His
                 30                  35                  40

GAC GTA CTT TTC AAA GGA AAT AAT TAC CAC AAT GCC ACT TCC GCC ATA     196
Asp Val Leu Phe Lys Gly Asn Asn Tyr His Asn Ala Thr Ser Ala Ile
             45                  50                  55

GTC GGG TCC CCC CAA TGG GGC AAC AAG ACT GCC ATG GCC GTG CCA TTC     244
Val Gly Ser Pro Gln Trp Gly Asn Lys Thr Ala Met Ala Val Pro Phe
         60                  65                  70

AAT TAT GGT GAC CTA GTT GTG TTC GAC GAT CCC ATT ACC TTA GAC AAC     292
Asn Tyr Gly Asp Leu Val Val Phe Asp Asp Pro Ile Thr Leu Asp Asn
     75                  80                  85

AAT CTG CAT TCA CCC CCA GTG GGT CGG GCG CAA GGG ATG TAC TTC TAT     340
Asn Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Met Tyr Phe Tyr
 90                  95                 100                 105

GAT CAA AAA AAT ACA TAC AAT GCT TGG CTA GGG TTC TCA TTT TTG TTC     388
Asp Gln Lys Asn Thr Tyr Asn Ala Trp Leu Gly Phe Ser Phe Leu Phe
                110                 115                 120

AAT TCA ACT AAG TAT GTT GGA ACC TTG AAC TTT GCT GGG GCT GAT CCA     436
Asn Ser Thr Lys Tyr Val Gly Thr Leu Asn Phe Ala Gly Ala Asp Pro
            125                 130                 135

TTG TTG AAC AAG ACT AGA GAC ATA TCA GTC ATT GGT GGA ACT GGT GAC     484
Leu Leu Asn Lys Thr Arg Asp Ile Ser Val Ile Gly Gly Thr Gly Asp
        140                 145                 150
```

```
TTT TTC ATG GCG AGA GGG GTT GCC ACT TTG ATG ACC GAT GCC TTT GAA      532
Phe Phe Met Ala Arg Gly Val Ala Thr Leu Met Thr Asp Ala Phe Glu
    155                 160                 165

GGG GAT GTG TAT TTC CGC CTT CGT GTC GAT ATT AAT TTG TAT GAA TGT      580
Gly Asp Val Tyr Phe Arg Leu Arg Val Asp Ile Asn Leu Tyr Glu Cys
170                 175                 180                 185

TGG TAAACAATTT AGCCGTATAT ATATATATAT ATGGCTATAC ATATTTCATA           633
Trp

GAATCCAGAT TTGCTGTTTC AAATGTGTGT TTCTTTAGTT GTGCCACCAA TAAAAAAATG    693

TACACATTAT TTAATAAATA TAATTATTTA ATGTGTTCAT TTTTGAAGTT AAATTTAAGT    753

TGTATTTATT TGATTATGTA TAAATTCTCT ATTAGTAAAA TAGTCAAAGT GACACATATT    813

CAAGACGACA TATGTAACTT TATTTCATAT CTTCAACAAG TTCAATAATG TCATATATAT    873

TGTACTATTG AAAAAAAAAA AAAAAAAA                                       901
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 186 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein: Forsythia intermedia PSD-Fil protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Met Val Ser Lys Thr Gln Ile Val Ala Leu Phe Leu Cys Phe Leu Thr
 1               5                  10                  15

Ser Thr Ser Ser Ala Thr Tyr Gly Arg Lys Pro Arg Pro Arg Arg Pro
            20                  25                  30

Cys Lys Glu Leu Val Phe Tyr Phe His Asp Val Leu Phe Lys Gly Asn
        35                  40                  45

Asn Tyr His Asn Ala Thr Ser Ala Ile Val Gly Ser Pro Gln Trp Gly
    50                  55                  60

Asn Lys Thr Ala Met Ala Val Pro Phe Asn Tyr Gly Asp Leu Val Val
65                  70                  75                  80

Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Pro Val
                85                  90                  95

Gly Arg Ala Gln Gly Met Tyr Phe Tyr Asp Gln Lys Asn Thr Tyr Asn
            100                 105                 110

Ala Trp Leu Gly Phe Ser Phe Leu Phe Asn Ser Thr Lys Tyr Val Gly
        115                 120                 125

Thr Leu Asn Phe Ala Gly Ala Asp Pro Leu Leu Asn Lys Thr Arg Asp
    130                 135                 140
```



```
Ile Ser Val Ile Gly Gly Thr Gly Asp Phe Phe Met Ala Arg Gly Val
145                 150                 155                 160

Ala Thr Leu Met Thr Asp Ala Phe Glu Gly Asp Val Tyr Phe Arg Leu
                165                 170                 175

Arg Val Asp Ile Asn Leu Tyr Glu Cys Trp
            180                 185
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 858 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia cDNA PSD-Fi2

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 19..573

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
AATTCGGCAC GAGGAAAA ATG GCA GCT AAA ACA CAA ACC ACA GCC CTT TTC      51
                    Met Ala Ala Lys Thr Gln Thr Thr Ala Leu Phe
                                190                 195

CTC TGC CTC CTC ATC TGC ATC TCC GCC GTG TAC GGC CAC AAA ACC AGG      99
Leu Cys Leu Leu Ile Cys Ile Ser Ala Val Tyr Gly His Lys Thr Arg
    200                 205                 210

TCT CGA CGC CCC TGT AAA GAG CTC GTT TTC TTC TTC CAC GAC ATC CTC     147
Ser Arg Arg Pro Cys Lys Glu Leu Val Phe Phe Phe His Asp Ile Leu
    215                 220                 225

TAC CTA GGA TAC AAT AGA AAC AAT GCC ACC GCT GTC ATA GTA GCC TCT     195
Tyr Leu Gly Tyr Asn Arg Asn Asn Ala Thr Ala Val Ile Val Ala Ser
230                 235                 240                 245

CCT CAA TGG GGA AAC AAG ACT GCC ATG GCT AAA CCT TTC AAT TTT GGT     243
Pro Gln Trp Gly Asn Lys Thr Ala Met Ala Lys Pro Phe Asn Phe Gly
                250                 255                 260

GAT TTG GTT GTG TTT GAT GAT CCC ATT ACC TTA GAC AAC AAC CTG CAT     291
Asp Leu Val Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His
            265                 270                 275

TCT CCT CCG GTC GGC CGG GCT CAG GGA ACT TAT TTC TAC GAT CAA TGG     339
Ser Pro Pro Val Gly Arg Ala Gln Gly Thr Tyr Phe Tyr Asp Gln Trp
        280                 285                 290
```

```
AGT ATT TAT GGT GCA TGG CTT GGA TTT TCA TTT TTG TTC AAT TCT ACT      387
Ser Ile Tyr Gly Ala Trp Leu Gly Phe Ser Phe Leu Phe Asn Ser Thr
    295                 300                 305

GAT TAT GTT GGA ACT CTA AAT TTT GCT GGA GCT GAT CCA TTG ATT AAC      435
Asp Tyr Val Gly Thr Leu Asn Phe Ala Gly Ala Asp Pro Leu Ile Asn
310                 315                 320                 325

AAA ACT AGG GAC ATT TCA GTA ATT GGA GGA ACT GGT GAT TTT TTC ATG      483
Lys Thr Arg Asp Ile Ser Val Ile Gly Gly Thr Gly Asp Phe Phe Met
                330                 335                 340

GCT AGA GGG GTA GCC ACT GTG TCG ACC GAT GCT TTT GAA GGG GAT GTT      531
Ala Arg Gly Val Ala Thr Val Ser Thr Asp Ala Phe Glu Gly Asp Val
            345                 350                 355

TAT TTC AGG CTT CGT GTT GAT ATT AGG TTG TAT GAG TGT TGG              573
Tyr Phe Arg Leu Arg Val Asp Ile Arg Leu Tyr Glu Cys Trp
        360                 365                 370

TAAATTTACC TTATTTTTCC ATTTTCTTGA GTTTGACTCG GATTTGACTA ATAATGTCTT    633

CTGTAATCCT TGTTTTTGAT CAATTTGTGG CGATTTTATC AATTAGTGAT TGTTTGGTTC    693

ATATTTTAAT CTGTTAAAAA AAATTGTGGT CAAAAGCCAA TAACCACAAC CGTAGGGAGT    753

TTTTTCCGTT AAGGGGAAAA AAAAGTATGT CCATGTGTTA CTACGTTTTC AATTTCATTC    813

AAAATTTGCT TTTCAATCAT CTTCTTCAAA AAAAAAAAAA AAAAA                    858
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 185 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia dirigent protein PSD-Fi2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
Met Ala Ala Lys Thr Gln Thr Thr Ala Leu Phe Leu Cys Leu Leu Ile
1               5                   10                  15

Cys Ile Ser Ala Val Tyr Gly His Lys Thr Arg Ser Arg Arg Pro Cys
            20                  25                  30

Lys Glu Leu Val Phe Phe Phe His Asp Ile Leu Tyr Leu Gly Tyr Asn
        35                  40                  45

Arg Asn Asn Ala Thr Ala Val Ile Val Ala Ser Pro Gln Trp Gly Asn
    50                  55                  60

Lys Thr Ala Met Ala Lys Pro Phe Asn Phe Gly Asp Leu Val Val Phe
65                  70                  75                  80

Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Pro Val Gly
                85                  90                  95
```



```
Arg Ala Gln Gly Thr Tyr Phe Tyr Asp Gln Trp Ser Ile Tyr Gly Ala
            100                 105                 110

Trp Leu Gly Phe Ser Phe Leu Phe Asn Ser Thr Asp Tyr Val Gly Thr
        115                 120                 125

Leu Asn Phe Ala Gly Ala Asp Pro Leu Ile Asn Lys Thr Arg Asp Ile
    130                 135                 140

Ser Val Ile Gly Gly Thr Gly Asp Phe Phe Met Ala Arg Gly Val Ala
145                 150                 155                 160

Thr Val Ser Thr Asp Ala Phe Glu Gly Asp Val Tyr Phe Arg Leu Arg
                165                 170                 175

Val Asp Ile Arg Leu Tyr Glu Cys Trp
            180                 185
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 948 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Tsuga heterophylla dirigent protein cDNA PSD-Th1

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 104..688

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
GGGCACCCTC TCTTGTTAAT TGAGCCCTTC TCCTCCTACT TCTCTTGTTA GTTCTTTGAT     60

CCCATATCTT CTTCTATAAT CACTTTAGTC TATAAGATTG TCA ATG GCA ATC AAG      115
                                                Met Ala Ile Lys

AAT CGT AAT AGA GCT GTG CAC TTG TGT TTT CTA TGG CTT CTA CTG TCC      163
Asn Arg Asn Arg Ala Val His Leu Cys Phe Leu Trp Leu Leu Leu Ser
190                 195                 200                 205

TCT GTG TTG TTG CAA ACA AGT GAT GGG AAA AGC TGG AAG AAG CAC CGA      211
Ser Val Leu Leu Gln Thr Ser Asp Gly Lys Ser Trp Lys Lys His Arg
                210                 215                 220

CTC CGA AAG CCT TGT AGG AAT CTG GTG TTG TAT TTC CAT GAT GTA ATC      259
Leu Arg Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp Val Ile
            225                 230                 235

TAC AAT GGC AGC AAC GCC AAG AAC GCT ACA TCC ACA CTT GTG GGT GCT      307
Tyr Asn Gly Ser Asn Ala Lys Asn Ala Thr Ser Thr Leu Val Gly Ala
        240                 245                 250
```

```
CCC CAC GGG TCT AAC CTC ACA CTT CTC GCT GGA AAA GAC AAC CAC TTT      355
Pro His Gly Ser Asn Leu Thr Leu Leu Ala Gly Lys Asp Asn His Phe
    255                 260                 265

GGA GAT CTG GCG GTG TTT GAC GAT CCC ATC ACT CTT GAC AAC AAT TTC      403
Gly Asp Leu Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Phe
270                 275                 280                 285

CAC TCT CCT CCG GTG GGC AGA GCT CAG GGA TTC TAC TTT TAT GAC ATG      451
His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met
                290                 295                 300

AAG AAC ACC TTC AGC TCC TGG CTT GGA TTC ACG TTT GTA CTC AAC TCT      499
Lys Asn Thr Phe Ser Ser Trp Leu Gly Phe Thr Phe Val Leu Asn Ser
            305                 310                 315

ACA GAT TAC AAA GGC ACC ATC ACG TTC TCT GGA GCC GAT CCA ATC CTT      547
Thr Asp Tyr Lys Gly Thr Ile Thr Phe Ser Gly Ala Asp Pro Ile Leu
        320                 325                 330

ACT AAA TAC AGA GAT ATA TCA GTG GTG GGA GGA ACT GGA GAT TTC ATA      595
Thr Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Ile
    335                 340                 345

ATG GCA AGA GGA ATC GCC ACA ATC TCC ACC GAT GCG TAT GAA GGC GAC      643
Met Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ala Tyr Glu Gly Asp
350                 355                 360                 365

GTT TAC TTC CGT CTC TGC GTG AAT ATC ACA CTC TAT GAG TGC TAC          688
Val Tyr Phe Arg Leu Cys Val Asn Ile Thr Leu Tyr Glu Cys Tyr
                370                 375                 380

TGAGTGCTAT AGGTCTATTT TCTCCTTCGA CTATCCATTT ATATGTTCAT TTTAGTTGAA    748

CTAGTGTTTT CTTGTGCGAG AGATATGCAC GAAGCTCTGA GATATTGTAG CGTGAAGTTC    808

CTTTAGCAGC CGAATAATGT ATTTCGATTT TGTCGAAGGC CATATCTAAT ATTGTCAAGG    868

GAAAATGCAG AATTCTATGT CGGTCAAGCA CTTTTATTTA AAAATAAAAG AAATATTGGT    928

TAAAAAAAAA AAAAAAAAAA                                                948
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 195 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Tsuga heterophylla dirigent protein PSD-Th1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
Met Ala Ile Lys Asn Arg Asn Arg Ala Val His Leu Cys Phe Leu Trp
1               5                   10                  15

Leu Leu Leu Ser Ser Val Leu Leu Gln Thr Ser Asp Gly Lys Ser Trp
            20                  25                  30
```



```
Lys Lys His Arg Leu Arg Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe
        35                  40                  45

His Asp Val Ile Tyr Asn Gly Ser Asn Ala Lys Asn Ala Thr Ser Thr
    50                  55                  60

Leu Val Gly Ala Pro His Gly Ser Asn Leu Thr Leu Leu Ala Gly Lys
65                  70                  75                  80

Asp Asn His Phe Gly Asp Leu Ala Val Phe Asp Asp Pro Ile Thr Leu
                85                  90                  95

Asp Asn Asn Phe His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr
            100                 105                 110

Phe Tyr Asp Met Lys Asn Thr Phe Ser Ser Trp Leu Gly Phe Thr Phe
        115                 120                 125

Val Leu Asn Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Ser Gly Ala
    130                 135                 140

Asp Pro Ile Leu Thr Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr
145                 150                 155                 160

Gly Asp Phe Ile Met Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ala
                165                 170                 175

Tyr Glu Gly Asp Val Tyr Phe Arg Leu Cys Val Asn Ile Thr Leu Tyr
            180                 185                 190

Glu Cys Tyr
        195
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 849 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Tsuga heterophylla dirigent protein PSD-Th2 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 71..625

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
GTTCTGTTCC AAATTCTAAT TAGCCTTCCA TTCATTCCAG GATCCCACTC TTCTTCCTTC    60

AAGATTGGCA ATG GCT ATC AAG AGT AAT AGG GCT GTG CGT TTC TGC TTT      109
           Met Ala Ile Lys Ser Asn Arg Ala Val Arg Phe Cys Phe
                       200                 205
```

```
GTA TGG CTT CTG TTG TTG CAA AGT GGT TTT GTA TTT CCA CTC CCA CAG      157
Val Trp Leu Leu Leu Leu Gln Ser Gly Phe Val Phe Pro Leu Pro Gln
    210                 215                 220

CCT TGT AGG AAT CTG GTT TTG TAT TTC CAC GAT GTA CTC TAC AAT GGC      205
Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp Val Leu Tyr Asn Gly
225                 230                 235                 240

TTC AAC GCC CAC AAC GCT ACA TCT ACA CTT GTG GGT GCT CCA CAG GGG      253
Phe Asn Ala His Asn Ala Thr Ser Thr Leu Val Gly Ala Pro Gln Gly
                245                 250                 255

GCT AAC CTC ACA CTT CTC GCT GGA AAA GAC AAC CAC TTT GGA GAT CTG      301
Ala Asn Leu Thr Leu Leu Ala Gly Lys Asp Asn His Phe Gly Asp Leu
            260                 265                 270

GCG GTG TTC GAC GAT CCC ATC ACT CTT GAC AAC AAT TTC CAG TCT CCT      349
Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Phe Gln Ser Pro
        275                 280                 285

CCG GTG GGC AGA GCT CAG GGA TTC TAC TTT TAT GAC ATG AAG AAC ACC      397
Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asn Thr
    290                 295                 300

TTC AGC TCC TGG CTT GGA TTC ACG TTT GTA CTC AAC TCT ACA GAT TAC      445
Phe Ser Ser Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp Tyr
305                 310                 315                 320

AAA GGC ACC ATC ACG TTC TCT GGA GCC GAT CCA ATC CTT ACT AAA TAC      493
Lys Gly Thr Ile Thr Phe Ser Gly Ala Asp Pro Ile Leu Thr Lys Tyr
                325                 330                 335

AGA GAT ATA TCA GTG GTG GGA GGA ACT GGA GAT TTC ATA ATG GCA AGA      541
Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Ile Met Ala Arg
            340                 345                 350

GGA ATC GCC ACA ATC TCC ACC GAT GCG TAT GAA GGA GAT GTT TAC TTC      589
Gly Ile Ala Thr Ile Ser Thr Asp Ala Tyr Glu Gly Asp Val Tyr Phe
        355                 360                 365

CGT CTC CGC GTC AAT ATC ACA CTC TAT GAA TGC TAC TGATATTATT           635
Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
    370                 375                 380

AAGTAGCTAC TGTTTCTCGT CTGGTCTCGC CATTTCGATG CTCTTTTTAA CATTAGTGCT    695

TTCCATAAAT TGTTGTAGCC TCTCAATAAA ACCCAGTAAA ATATTTCTTC TGTTTATTTA    755

GCAGCTTCCA AATCATTGTA TTAGTATTTT ATATTATTTG GATTTTATAC AAGTCCATAA    815

AATATTTCTT CAGCTAAAAA AAAAAAAAAA AAAA                                849
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 185 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Tsuga heterophylla dirigent protein translated from PSD-Th2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
Met Ala Ile Lys Ser Asn Arg Ala Val Arg Phe Cys Phe Val Trp Leu
1               5                   10                  15

Leu Leu Leu Gln Ser Gly Phe Val Phe Pro Leu Pro Gln Pro Cys Arg
            20                  25                  30

Asn Leu Val Leu Tyr Phe His Asp Val Leu Tyr Asn Gly Phe Asn Ala
        35                  40                  45

His Asn Ala Thr Ser Thr Leu Val Gly Ala Pro Gln Gly Ala Asn Leu
    50                  55                  60

Thr Leu Leu Ala Gly Lys Asp Asn His Phe Gly Asp Leu Ala Val Phe
65                  70                  75                  80

Asp Asp Pro Ile Thr Leu Asp Asn Asn Phe Gln Ser Pro Pro Val Gly
                85                  90                  95

Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asn Thr Phe Ser Ser
            100                 105                 110

Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp Tyr Lys Gly Thr
        115                 120                 125

Ile Thr Phe Ser Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg Asp Ile
    130                 135                 140

Ser Val Val Gly Gly Thr Gly Asp Phe Ile Met Ala Arg Gly Ile Ala
145                 150                 155                 160

Thr Ile Ser Thr Asp Ala Tyr Glu Gly Asp Val Tyr Phe Arg Leu Arg
                165                 170                 175

Val Asn Ile Thr Leu Tyr Glu Cys Tyr
            180                 185
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 873 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp1 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 25..591

```
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTTGCACGAG GGATTTCAAG AGAT ATG AGT AGA ATA GCA TTT CAT TTG TGC        51
                           Met Ser Arg Ile Ala Phe His Leu Cys
                                           190

TTC ATG GGG CTT CTG CTC TCT TCC ACG GTG CTC AGA AAT GTA GAT GGG       99
Phe Met Gly Leu Leu Leu Ser Ser Thr Val Leu Arg Asn Val Asp Gly
195                 200                 205                 210

CAT GCA TGG AAG AGG CAA CTT CCA ATG CCA TGT AAG AAT TTG GTG CTC      147
His Ala Trp Lys Arg Gln Leu Pro Met Pro Cys Lys Asn Leu Val Leu
                215                 220                 225

TAC TTT CAT GAT ATA CTC TAC AAT GGC AAA AAC ATT CAC AAT GCA ACT      195
Tyr Phe His Asp Ile Leu Tyr Asn Gly Lys Asn Ile His Asn Ala Thr
            230                 235                 240

GCT GCG CTG GTT GCA GCT CCT GCG TGG GGC AAT CTC ACT ACT TTC GCT      243
Ala Ala Leu Val Ala Ala Pro Ala Trp Gly Asn Leu Thr Thr Phe Ala
        245                 250                 255

GAA CCT TTC AAG TTT GGA GAT GTG GTT GTG TTT GAC GAT CCC ATT ACT      291
Glu Pro Phe Lys Phe Gly Asp Val Val Val Phe Asp Asp Pro Ile Thr
    260                 265                 270

CTC GAC AAC AAT CTT CAC TCT CCT CCT GTG GGA AGA GCG CAG GGA TTT      339
Leu Asp Asn Asn Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe
275                 280                 285                 290

TAT TTG TAC AAC ATG AAG ACT ACT TAC AAT GCT TGG TTG GGG TTC ACA      387
Tyr Leu Tyr Asn Met Lys Thr Thr Tyr Asn Ala Trp Leu Gly Phe Thr
                295                 300                 305

TTT GTG CTG AAT TCG ACA GAT TAT AAG GGC ACA ATC ACC TTC AAT GGC      435
Phe Val Leu Asn Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Asn Gly
            310                 315                 320

GCC GAC CCC CCG CTG GTT AAG TAC AGA GAT ATA TCC GTT GTT GGC GGT      483
Ala Asp Pro Pro Leu Val Lys Tyr Arg Asp Ile Ser Val Val Gly Gly
        325                 330                 335

ACG GGT GAT TTC TTG ATG GCG AGA GGA ATT GCC ACC CTT TCT ACT GAT      531
Thr Gly Asp Phe Leu Met Ala Arg Gly Ile Ala Thr Leu Ser Thr Asp
    340                 345                 350

GCA ATC GAG GGA AAT GTT TAT TTC CGA CTC AGG GTT AAC ATC ACA CTC      579
Ala Ile Glu Gly Asn Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu
355                 360                 365                 370

TAC GAG TGT TAC TGATGATTAC TAACTAAATG GAGAGTCTTT GTTTAGAGAA           631
Tyr Glu Cys Tyr

TAGTGTGTTG GGCTGTTTAC TTAAAGTCGA CGTTCTATGC AGTTGAAGTC TTTGTTTAGA     691

TGAATGCAAT GGTGGGTTTT CTTTCCTCGT GAGGGTTAAC ATCACACTCT ACGAGTGTTA     751

CTGATAATTG TTAAGTATTT GGAGAGTCTT GTAAGTTGAG AATAATGTAT TTTGGCTGTT     811
```



```
TATTTTGAGT CGAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA     871

AA                                                                    873
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 189 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

```
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met Ser Arg Ile Ala Phe His Leu Cys Phe Met Gly Leu Leu Leu Ser
1               5                   10                  15

Ser Thr Val Leu Arg Asn Val Asp Gly His Ala Trp Lys Arg Gln Leu
            20                  25                  30

Pro Met Pro Cys Lys Asn Leu Val Leu Tyr Phe His Asp Ile Leu Tyr
        35                  40                  45

Asn Gly Lys Asn Ile His Asn Ala Thr Ala Ala Leu Val Ala Ala Pro
    50                  55                  60

Ala Trp Gly Asn Leu Thr Thr Phe Ala Glu Pro Phe Lys Phe Gly Asp
65                  70                  75                  80

Val Val Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser
                85                  90                  95

Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Leu Tyr Asn Met Lys Thr
            100                 105                 110

Thr Tyr Asn Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp
        115                 120                 125

Tyr Lys Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Pro Leu Val Lys
    130                 135                 140

Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala
145                 150                 155                 160

Arg Gly Ile Ala Thr Leu Ser Thr Asp Ala Ile Glu Gly Asn Val Tyr
                165                 170                 175

Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
            180                 185
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 867 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp2 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 80..655

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
GCAATATTGT GCTGGTTCAG TAATCTATGT CTTGTTGACC TGTAGTGTAT ACCCAAACAT      60

TTCTCCTTCT TTTGCAAAA ATG GCA ATG AAG GCT GCA AAA TTT CTG CAT TTC      112
                     Met Ala Met Lys Ala Ala Lys Phe Leu His Phe
                     190                 195                 200

TTA TTT ATC TGG CTT CTA GTC TGC ACT GTG TTG CTC AAA TCT GCA GAC      160
Leu Phe Ile Trp Leu Leu Val Cys Thr Val Leu Leu Lys Ser Ala Asp
                205                 210                 215

TGT CAT AGA TGG AAG AAG AAA ATT CCA GAG CCA TGT AAG AAT CTG GTA      208
Cys His Arg Trp Lys Lys Lys Ile Pro Glu Pro Cys Lys Asn Leu Val
            220                 225                 230

TTG TAC TTT CAT GAT ATC CTC TAC AAT GGA TCC AAC AAA CAC AAT GCA      256
Leu Tyr Phe His Asp Ile Leu Tyr Asn Gly Ser Asn Lys His Asn Ala
        235                 240                 245

ACA TCT GCA ATT GTT GGA GCA CCC AAA GGA GCC AAT CTC ACT ATT TTG      304
Thr Ser Ala Ile Val Gly Ala Pro Lys Gly Ala Asn Leu Thr Ile Leu
    250                 255                 260

ACT GGT AAC AAC CAT TTT GGA GAT GTG GTT GTG TTT GAT GAT CCT ATT      352
Thr Gly Asn Asn His Phe Gly Asp Val Val Val Phe Asp Asp Pro Ile
265                 270                 275                 280

ACT CTT GAC AAC AAT CTT CAC TCT ACT CCT GTG GGA AGA GCT CAG GGC      400
Thr Leu Asp Asn Asn Leu His Ser Thr Pro Val Gly Arg Ala Gln Gly
                285                 290                 295

TTT TAT TTC TAT GAC ATG AAG AAT ACA TTC AAT TCT TGG CTT GGG TTT      448
Phe Tyr Phe Tyr Asp Met Lys Asn Thr Phe Asn Ser Trp Leu Gly Phe
            300                 305                 310

ACA TTT GTG TTG AAT TCA ACA AAT TAT AAG GGC ACC ATC ACC TTC AAT      496
Thr Phe Val Leu Asn Ser Thr Asn Tyr Lys Gly Thr Ile Thr Phe Asn
        315                 320                 325

GGG GCT GAC CCA ATT CTG ACT AAG TAC AGA GAT ATA TCT GTT GTG GGT      544
Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg Asp Ile Ser Val Val Gly
    330                 335                 340

GGT ACG GGT GAT TTC TTG ATG GCC AGA GGA ATC GCC ACC ATT TCT ACT      592
Gly Thr Gly Asp Phe Leu Met Ala Arg Gly Ile Ala Thr Ile Ser Thr
345                 350                 355                 360
```



```
GAT GCA TAC GAG GGA GAT GTT TAT TTC CGT CTT AGG GTG AAT ATC ACT      640
Asp Ala Tyr Glu Gly Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr
                365                 370                 375

CTC TAT GAG TGT TAC TGATTCGAAT TTGATTTCCT GTTCTAATCT CTAATTTGAG      695
Leu Tyr Glu Cys Tyr
            380

AGGATGAACA TTCAATAAAC TTTATAGAAG CATATATAAA TAGGTGCAGG AAAATAAGAG      755

GTAAGGGATG AGATTATTTC AGCCTCATAT CTTATTCTGC ATCAGTTTTG TATGCTCATT      815

TGTTTAATAA AATTTGACCA GTTTCATCAT GTTGAAAAAA AAAAAAAAAA AA            867
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
Met Ala Met Lys Ala Ala Lys Phe Leu His Phe Leu Phe Ile Trp Leu
  1               5                  10                  15

Leu Val Cys Thr Val Leu Leu Lys Ser Ala Asp Cys His Arg Trp Lys
             20                  25                  30

Lys Lys Ile Pro Glu Pro Cys Lys Asn Leu Val Leu Tyr Phe His Asp
         35                  40                  45

Ile Leu Tyr Asn Gly Ser Asn Lys His Asn Ala Thr Ser Ala Ile Val
     50                  55                  60

Gly Ala Pro Lys Gly Ala Asn Leu Thr Ile Leu Thr Gly Asn Asn His
 65                  70                  75                  80

Phe Gly Asp Val Val Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn
                 85                  90                  95

Leu His Ser Thr Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp
            100                 105                 110

Met Lys Asn Thr Phe Asn Ser Trp Leu Gly Phe Thr Phe Val Leu Asn
        115                 120                 125

Ser Thr Asn Tyr Lys Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile
    130                 135                 140

Leu Thr Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe
145                 150                 155                 160

Leu Met Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ala Tyr Glu Gly
                165                 170                 175
```

```
Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
        180                 185                 190
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp3 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 94..669

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
CGTAGGAAAT ATCTCAGAGG GAGCCGAAAA TTGAGATAAT TGTTGTACGA AATATATAAA      60

AGATTAGATT CAGAGGAATT TGCAGATGTT GTT GTA TCT AAA ACA GCT GCT AGA     114
                                     Val Ser Lys Thr Ala Ala Arg
                                                 195

GTT CTG CAT TTA TGC TTT CTA TGG CTT CTA GTA TCT GCA ATC TTC ATA     162
Val Leu His Leu Cys Phe Leu Trp Leu Leu Val Ser Ala Ile Phe Ile
200                 205                 210                 215

AAA TCT GCA GAT TGC CGT AGC TGG AAA AAG AAG CTT CCA AAG CCC TGT     210
Lys Ser Ala Asp Cys Arg Ser Trp Lys Lys Lys Leu Pro Lys Pro Cys
                220                 225                 230

AGA AAT CTT GTG TTA TAT TTT CAT GAT ATA ATC TAC AAT GGC AAA AAT     258
Arg Asn Leu Val Leu Tyr Phe His Asp Ile Ile Tyr Asn Gly Lys Asn
            235                 240                 245

GCA GAG AAT GCA ACA TCT GCA CTT GTT TCA GCC CCT CAA GGA GCT AAT     306
Ala Glu Asn Ala Thr Ser Ala Leu Val Ser Ala Pro Gln Gly Ala Asn
        250                 255                 260

CTC ACC ATT ATG ACT GGT AAT AAC CAT TTT GGG AAT CTT GCA GTG TTT     354
Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly Asn Leu Ala Val Phe
    265                 270                 275

GAT GAT CCT ATT ACT CTT GAC AAC AAT CTT CAC TCT CCT CCT GTT GGA     402
Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Pro Val Gly
280                 285                 290                 295

AGA GCT CAG GGC TTT TAC TTC TAT GAC ATG AAG AAC ACC TTC AGT GCC     450
Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asn Thr Phe Ser Ala
                300                 305                 310
```



```
TGG CTT GGC TTC ACA TTT GTG CTC AAT TCA ACT GAT CAC AAG GGC TCC     498
Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp His Lys Gly Ser
            315                 320                 325

ATT ACT TTC AAT GGA GCA GAT CCC ATC TTA ACA AAG TAC AGA GAC ATA     546
Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg Asp Ile
        330                 335                 340

TCT GTT GTG GGT GGA ACA GGG GAT TTC TTG ATG GCA AGA GGA ATT GCT     594
Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg Gly Ile Ala
    345                 350                 355

ACC ATT TCT ACT GAC TCA TAT GAG GGA GAT GTT TAT TTC AGG CTT AGG     642
Thr Ile Ser Thr Asp Ser Tyr Glu Gly Asp Val Tyr Phe Arg Leu Arg
360                 365                 370                 375

GTC AAT ATC ACA CTC TAT GAG TGT TAC TGAACAAATT CCTTGCTCTG           689
Val Asn Ile Thr Leu Tyr Glu Cys Tyr
                380

TATTTCTAGT TTTTGGGACC TTTTAAAGAT AGTTGTTTAC TTCAATGTCT CTATATGTAA    749

TAACACTGTG TGAAGATTAT ATACGATGGA CTATAGAAAC TATGTTGAAT TCTGTTCTGT    809

AGCTAATTTA TGTATATGAT CCACTCATAT CTCTTAATAT GATACCGATT TGTAATTATC    869

CCAGATAAAG TATGTCATGT GCTTTGACAA AAAAAAAAAA AAAAA                    914
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
Val Ser Lys Thr Ala Ala Arg Val Leu His Leu Cys Phe Leu Trp Leu
 1               5                  10                  15

Leu Val Ser Ala Ile Phe Ile Lys Ser Ala Asp Cys Arg Ser Trp Lys
            20                  25                  30

Lys Lys Leu Pro Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp
        35                  40                  45

Ile Ile Tyr Asn Gly Lys Asn Ala Glu Asn Ala Thr Ser Ala Leu Val
    50                  55                  60

Ser Ala Pro Gln Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His
65                  70                  75                  80

Phe Gly Asn Leu Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn
                85                  90                  95

Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp
            100                 105                 110
```

```
Met Lys Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn
        115                 120                 125

Ser Thr Asp His Lys Gly Ser Ile Thr Phe Asn Gly Ala Asp Pro Ile
    130                 135                 140

Leu Thr Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe
145                 150                 155                 160

Leu Met Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ser Tyr Glu Gly
                165                 170                 175

Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
            180                 185                 190
```

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 704 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp4 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..416

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
AG AAT GCC CAC AAT GCA ACA TCT GCA CTT GTT GCA GCC CCT GAG GGA      47
   Asn Ala His Asn Ala Thr Ser Ala Leu Val Ala Ala Pro Glu Gly
       195                 200                 205

GCC AAT CTC ACC ATT ATG ACT GGT AAT AAC CAT TTT GGG AAT ATT GCT     95
Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly Asn Ile Ala
        210                 215                 220

GTG TTT GAT GAT CCT ATT ACT CTT GAC AAC AAT CTT CAC TCT CCT TCT    143
Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Ser
    225                 230                 235

GTT GGA AGA GCT CAG GGC TTT TAC TTC TAT GAC ATG AAG GAT ACC TTC    191
Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asp Thr Phe
240                 245                 250                 255

AAT GCT TGG CTT GGT TTT ACA TTT GTG CTG AAT TCA ACT GAT CAC AAG    239
Asn Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp His Lys
                260                 265                 270

GGC ACC ATT ACT TTC AAT GGA GCA GAT CCA ATC CTG ACC AAG TAC AGA    287
Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg
            275                 280                 285
```



```
GAT ATA TCT GTT GTG GGT GGA ACA GGG GAT TTC TTG ATG GCC AGA GGA    335
Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg Gly
        290                 295                 300

ATT GCC ACC ATT TCT ACT GAT TCA TAT GAG GGA GAT GTT TAT TTC AGG    383
Ile Ala Thr Ile Ser Thr Asp Ser Tyr Glu Gly Asp Val Tyr Phe Arg
    305                 310                 315

CTT AGG GTC AAT ATC ACA CTC TAT GAG TGT TAC TAAAAATGAA TTTCCTCTGT  436
Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
320                 325                 330

ATTACTAGCT TATAGGAGTC ATTCCCTGGT TCAATGTCTA GGGCATGGAA TAAAAGAATT  496

TGAAGATGGT TTTGAAATAT GGAGCATGTA TTCTAATTTG AAGAGCCCTC AAGGAAGTGC  556

ATTTTACAGA GTTTAGTTTT GCCCTCTAGA ATATTATGTT TTCAAAATGC TCTATGAAAG  616

TCATATGATG TATGGAGTAC CATTTGGAAT AATTAAAGCA AGCATATTTT ATTAAAAAAA  676

AAAAAAAAAA AAAAAAAAAA AAAAAAA                                      704
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 138 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
Asn Ala His Asn Ala Thr Ser Ala Leu Val Ala Ala Pro Glu Gly Ala
  1               5                  10                  15

Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly Asn Ile Ala Val
             20                  25                  30

Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Ser Val
         35                  40                  45

Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asp Thr Phe Asn
     50                  55                  60

Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp His Lys Gly
 65                  70                  75                  80

Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg Asp
                 85                  90                  95

Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg Gly Ile
            100                 105                 110

Ala Thr Ile Ser Thr Asp Ser Tyr Glu Gly Asp Val Tyr Phe Arg Leu
        115                 120                 125
```

```
Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
    130                 135
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 820 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp5 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 43..612

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
GTCTAATTGA GAGAAAATTC CAATAATTTT TTACCAATAG CA ATG AAA GCC ATT      54
                                               Met Lys Ala Ile
                                                   140

AGA GTT CTG CAT TTA TGC TTT CTA TGT CTT CTA GTG TCT GCA ATC TTG    102
Arg Val Leu His Leu Cys Phe Leu Cys Leu Leu Val Ser Ala Ile Leu
        145                 150                 155

CTA AAA TCT GCA GAT TGC CAT AGC TGG AAA AAG AAG CTT CCA AAG CCC    150
Leu Lys Ser Ala Asp Cys His Ser Trp Lys Lys Lys Leu Pro Lys Pro
    160                 165                 170

TGC AAG AAT CTT GTG TTA TAT TTC CAT GAT ATA ATC TAC AAT GGC AAA    198
Cys Lys Asn Leu Val Leu Tyr Phe His Asp Ile Ile Tyr Asn Gly Lys
175                 180                 185                 190

AAT GCA GAG AAT GCA ACA TCT GCA CTT GTT GCA GCC CCT GAG GGA GCC    246
Asn Ala Glu Asn Ala Thr Ser Ala Leu Val Ala Ala Pro Glu Gly Ala
                195                 200                 205

AAT CTC ACC ATT ATG ACT GGT AAT AAC CAT TTT GGG AAT CTT GCT GTG    294
Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly Asn Leu Ala Val
            210                 215                 220

TTT GAT GAT CCT ATT ACT CTT GAC AAC AAT CTC CAC TCT CCT CCT GTG    342
Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Pro Val
        225                 230                 235

GGA AGA GCT CAG GGA TTT TAC TTC TAT GAC ATG AAG AAC ACC TTC AGT    390
Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asn Thr Phe Ser
    240                 245                 250

GCT TGG CTT GGC TTC ACA TTT GTG CTG AAT TCA ACT GAT CAC AAG GGC    438
Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp His Lys Gly
255                 260                 265                 270
```



```
ACC ATT ACT TTC AAT GGA GCA GAC CCA ATC CTG ACC AAG TAC AGA GAC    486
Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg Asp
                275                 280                 285

ATA TCT GTT GTG GGT GGA ACA GGG GAT TTC TTG ATG GCC AGA GGA ATT    534
Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg Gly Ile
            290                 295                 300

GCC ACC ATT TCT ACT GAT TCA TAT GAG GGA GAA GTT TAT TTC AGG CTT    582
Ala Thr Ile Ser Thr Asp Ser Tyr Glu Gly Glu Val Tyr Phe Arg Leu
        305                 310                 315

AGG GTC AAT ATC ACA CTC TAT GAG TGT TAC TGAGCAAATG CCTGTCTTCT      632
Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
    320                 325

TCCTCTGTAG TTCTTGTTTT GGGTGCCTTT GAGGAATAGT TCTTGGCTTC AATGTCTCTG  692

TATGTAGTAA CATGGTCAAT GGAGTCTATT TTGAAGATTA TGAAGATATA GTCTCTATAT  752

ATATATATAT TGAAGAGAAT GAGATCTGTT TTAGGTAGCT CTTTTCATTC AAAAAAAAAA  812

AAAAAAAA                                                           820
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 190 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
Met Lys Ala Ile Arg Val Leu His Leu Cys Phe Leu Cys Leu Leu Val
1               5                   10                  15

Ser Ala Ile Leu Leu Lys Ser Ala Asp Cys His Ser Trp Lys Lys Lys
            20                  25                  30

Leu Pro Lys Pro Cys Lys Asn Leu Val Leu Tyr Phe His Asp Ile Ile
        35                  40                  45

Tyr Asn Gly Lys Asn Ala Glu Asn Ala Thr Ser Ala Leu Val Ala Ala
    50                  55                  60

Pro Glu Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly
65                  70                  75                  80

Asn Leu Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His
                85                  90                  95

Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys
            100                 105                 110

Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr
        115                 120                 125
```

```
Asp His Lys Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr
    130                 135                 140

Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met
145                 150                 155                 160

Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ser Tyr Glu Gly Glu Val
                165                 170                 175

Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
            180                 185                 190
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1013 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp6 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 47..616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
CTCAGTCTAA TTGAGAGAAA ATTCCAATAA TTTTTTCCCA ATAGCA ATG AAA GCC        55
                                                   Met Lys Ala

ATT AGA GTT CTG CAA TTA TGC TTT CTA TGG CTT CTA GTA TCT GCA ATC      103
Ile Arg Val Leu Gln Leu Cys Phe Leu Trp Leu Leu Val Ser Ala Ile
    195                 200                 205

TTG CTA AAA TCT GCA GAT TGC CAT AGC TGG AAA AAG AAG CTT CCA AAG      151
Leu Leu Lys Ser Ala Asp Cys His Ser Trp Lys Lys Lys Leu Pro Lys
210                 215                 220                 225

CCC TGC AAG AAT CTT GTG TTA TAT TTC CAT GAT ATA ATC TAC AAT GGC      199
Pro Cys Lys Asn Leu Val Leu Tyr Phe His Asp Ile Ile Tyr Asn Gly
                230                 235                 240

AAA AAT GCA GAG AAT GCA ACA TCT GCA CTT GTT GCA GCC CCT GAG GGA      247
Lys Asn Ala Glu Asn Ala Thr Ser Ala Leu Val Ala Ala Pro Glu Gly
            245                 250                 255

GCC AAT CTC ACC ATT ATG ACT GGT AAT AAC CAT TTT GGG AAT CTT GCT      295
Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly Asn Leu Ala
        260                 265                 270
```



```
GTG TTT GAT GAT CCT ATT ACT CTT GAC AAC AAT CTC CAC TCT CCT CCT      343
Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His Ser Pro Pro
    275                 280                 285

GTG GGA AGA GCT CAG GGC TTT TAC TTC TAT GAC ATG AAG AAC ACC TTC      391
Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asn Thr Phe
290                 295                 300                 305

AGT GCT TGG CTT GGC TTC ACA TTT GTG CTG AAT TCA ACT GAT CAC AAG      439
Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp His Lys
                310                 315                 320

GGC ACC ATT ACT TTC AAT GGA GCA GAC CCA ATC CTG ACC AAG TAC AGA      487
Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr Lys Tyr Arg
            325                 330                 335

GAT ATA TCT GTT GTG GGT GGA ACA GGG GAT TTC TTG ATG GCC AGA GGA      535
Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg Gly
        340                 345                 350

ATT GCC ACC ATT TCT ACT GAT TCA TAT GAG GGA GAT GTT TAT TTC AGG      583
Ile Ala Thr Ile Ser Thr Asp Ser Tyr Glu Gly Asp Val Tyr Phe Arg
    355                 360                 365

CTT AGG GTC AAT ATC ACA CTC TAT AAG TGT TAC TGAGCAAATG CCTGTCTTCT    636
Leu Arg Val Asn Ile Thr Leu Tyr Lys Cys Tyr
370                 375                 380

TCCTCTGTAG TTCTTGTTTT GGGTGCCTTT GAGGAATAGT TCTTGGCTTC AATGTCTCTG    696

TATGTAGTAA CATGGTCAAT GGAGTCTATT TTGAAGATTA TGAAGATATA GTCTCTCTAT    756

ATATATATAT TGAAGAGAAT GAGATCTGTT TTAGGTAGCT CTTTTCATTC ATATATATGG    816

GTTAACTTGG ATTTCATGTT TGGTTCAAAG ATCAGTTATG GAGGATTTCC TTTTAGTGGT    876

TTTATGGGAT TTTTGACATA TTAGATTACT TTCATCTCAA ATATATGTTA AATCAGTTAT    936

ATATGAAACT AATCATATAT AAGTTCAGAA ATATCAGAAC AACCATTTTA TGGAAAAAAA    996

AAAAAAAAAA AAAAAAA                                                  1013
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 190 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
Met Lys Ala Ile Arg Val Leu Gln Leu Cys Phe Leu Trp Leu Leu Val
1               5                   10                  15

Ser Ala Ile Leu Leu Lys Ser Ala Asp Cys His Ser Trp Lys Lys Lys
            20                  25                  30
```

```
Leu Pro Lys Pro Cys Lys Asn Leu Val Leu Tyr Phe His Asp Ile Ile
         35                  40                  45

Tyr Asn Gly Lys Asn Ala Glu Asn Ala Thr Ser Ala Leu Val Ala Ala
     50                  55                  60

Pro Glu Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Asn His Phe Gly
 65                  70                  75                  80

Asn Leu Ala Val Phe Asp Asp Pro Ile Thr Leu Asp Asn Asn Leu His
                 85                  90                  95

Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys
            100                 105                 110

Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr
        115                 120                 125

Asp His Lys Gly Thr Ile Thr Phe Asn Gly Ala Asp Pro Ile Leu Thr
    130                 135                 140

Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met
145                 150                 155                 160

Ala Arg Gly Ile Ala Thr Ile Ser Thr Asp Ser Tyr Glu Gly Asp Val
                165                 170                 175

Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Lys Cys Tyr
            180                 185                 190
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 913 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp7 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 77..652

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
GCAAGCTCAA ATACCCGACT TCTTTCTCTA CTTCAGAGCT CTTCTTTCTT CAAACATTTT      60

TGATATATTT TGCACA ATG GCA ATC TGG AAT GGA AGA GTT CTG AAT TTG         109
                  Met Ala Ile Trp Asn Gly Arg Val Leu Asn Leu
                                      195                 200

TGC ATT CTG TGG CTT CTG GTC TCC ATA GTT TTG CTG AAT GGT ATA GAT       157
Cys Ile Leu Trp Leu Leu Val Ser Ile Val Leu Leu Asn Gly Ile Asp
            205                 210                 215
```



```
TGC CAT AGT AGA AAA AAG AAG CTT CCA AAG CCA TGT AGG AAT CTT GTT       205
Cys His Ser Arg Lys Lys Lys Leu Pro Lys Pro Cys Arg Asn Leu Val
        220                 225                 230

TTG TAT TTT CAT GAT ATT ATC TAC AAT GGT AAA AAT GCA GGC AAT GCA       253
Leu Tyr Phe His Asp Ile Ile Tyr Asn Gly Lys Asn Ala Gly Asn Ala
    235                 240                 245

ACA TCT ACG CTT GTT GCA GCC CCT CAA GGA GCT AAT CTC ACC ATT ATG       301
Thr Ser Thr Leu Val Ala Ala Pro Gln Gly Ala Asn Leu Thr Ile Met
250                 255                 260                 265

ACT GGC AAT TAC CAT TTT GGA GAT CTG TCT GTG TTT GAT GAT CCT ATT       349
Thr Gly Asn Tyr His Phe Gly Asp Leu Ser Val Phe Asp Asp Pro Ile
                270                 275                 280

ACT GTT GAC AAC AAT CTT CAT TCT CCT CCT GTG GGA AGA GCT CAG GGC       397
Thr Val Asp Asn Asn Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly
            285                 290                 295

TTT TAC TTC TAT GAC ATG AAG AAT ACA TTC AGT GCT TGG CTT GGG TTC       445
Phe Tyr Phe Tyr Asp Met Lys Asn Thr Phe Ser Ala Trp Leu Gly Phe
        300                 305                 310

ACA TTT GTG CTG AAC TCA ACA GAT TAT AAA GGC ACT ATT ACT TTC GGT       493
Thr Phe Val Leu Asn Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Gly
    315                 320                 325

GGA GCA GAC CCA ATT TTG GCT AAG TAC AGA GAT ATA TCT GTT GTG GGT       541
Gly Ala Asp Pro Ile Leu Ala Lys Tyr Arg Asp Ile Ser Val Val Gly
330                 335                 340                 345

GGT ACT GGA GAT TTC TTG ATG GCA AGA GGA ATT GCT ACA ATC GAT ACT       589
Gly Thr Gly Asp Phe Leu Met Ala Arg Gly Ile Ala Thr Ile Asp Thr
                350                 355                 360

GAT GCA TAT GAG GGA GAT GTT TAT TTC AGG CTA AGG GTG AAT ATC ACA       637
Asp Ala Tyr Glu Gly Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr
            365                 370                 375

CTC TAT GAG TGT TAC TGATCCATGG GTATTCTATG TAGAATAGCT CAATCTGATA       692
Leu Tyr Glu Cys Tyr
        380

TGGCTATATT ATTTTGAGAG CATAGGTAGT TAAGTTTTAT AACTAAGTAG TGAACCATGA     752

GATCATTGAA AACTTGGGTG CTCATGCAGA GTTTTCATAT TTTCTAAATA AGTCTGCTCG     812

ACTATTACAT TTATGGATTG TTGAGAATTG TGTCGCTTAT TACTTTATGA ATAAGCTATT     872

TTAAACAAAG TTTTCACAAG TTTAAAAAAA AAAAAAAAAA A                         913
```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
Met Ala Ile Trp Asn Gly Arg Val Leu Asn Leu Cys Ile Leu Trp Leu
 1               5                  10                  15

Leu Val Ser Ile Val Leu Leu Asn Gly Ile Asp Cys His Ser Arg Lys
            20                  25                  30

Lys Lys Leu Pro Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp
        35                  40                  45

Ile Ile Tyr Asn Gly Lys Asn Ala Gly Asn Ala Thr Ser Thr Leu Val
    50                  55                  60

Ala Ala Pro Gln Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Tyr His
65                  70                  75                  80

Phe Gly Asp Leu Ser Val Phe Asp Asp Pro Ile Thr Val Asp Asn Asn
                85                  90                  95

Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp
            100                 105                 110

Met Lys Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn
        115                 120                 125

Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Gly Gly Ala Asp Pro Ile
    130                 135                 140

Leu Ala Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe
145                 150                 155                 160

Leu Met Ala Arg Gly Ile Ala Thr Ile Asp Thr Asp Ala Tyr Glu Gly
                165                 170                 175

Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
            180                 185                 190
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 890 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata dirigent protein PSD-Tp8 cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 44..619

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
CAGAGCTCTT CTTTCTTCAA ACATTTTTGA TATATTTTGC ACA ATG GCA ATC TGG      55
                                                Met Ala Ile Trp
                                                        195

AAT GGA AGA GTT CTG AAT TTG TGC ATT CTG TGG CTT CTG GTC TCC ATA     103
Asn Gly Arg Val Leu Asn Leu Cys Ile Leu Trp Leu Leu Val Ser Ile
            200                 205                 210

GTT TTG CTG AAT GGT ATA GAT TGC CAT AGT AGA AAA AAG AAG CTT CCA     151
Val Leu Leu Asn Gly Ile Asp Cys His Ser Arg Lys Lys Lys Leu Pro
        215                 220                 225

AAG CCA TGT AGG AAT CTT GTT TTG TAT TTT CAT GAT ATT ATC TAC AAT     199
Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp Ile Ile Tyr Asn
    230                 235                 240

GGT AAA AAT GCA GGC AAT GCA ACA TCT ACG CTT GTT GCA GCC CCT CAA     247
Gly Lys Asn Ala Gly Asn Ala Thr Ser Thr Leu Val Ala Ala Pro Gln
245                 250                 255                 260

GGA GCT AAT CTC ACC ATT ATG ACT GGC AAT TAC CAT TTT GGA GAT CTG     295
Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Tyr His Phe Gly Asp Leu
                265                 270                 275

GCT GTG TTT GAT GAT CCT ATT ACT GTT GAC AAC AAT CTT CAT TCT CCT     343
Ala Val Phe Asp Asp Pro Ile Thr Val Asp Asn Asn Leu His Ser Pro
            280                 285                 290

CCT GTG GGA AGA GCT CAG GGC TTT TAC TTC TAT GAC ATG AAG AAT ACA     391
Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp Met Lys Asn Thr
        295                 300                 305

TTC AGT GCT TGG CTT GGG TTC ACA TTT GTG CTG AAC TCA ACA GAT TAT     439
Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn Ser Thr Asp Tyr
    310                 315                 320

AAA GGC ACT ATT ACT TTC GGT GGA GCA GAC CCA ATT TTG GCT AAG TAC     487
Lys Gly Thr Ile Thr Phe Gly Gly Ala Asp Pro Ile Leu Ala Lys Tyr
325                 330                 335                 340

AGA GAT ATA TCT GTT GTG GGT GGT ACT GGA GAT TTC TTG ATG GCA AGA     535
Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe Leu Met Ala Arg
                345                 350                 355

GGA ATT GCT ACA ATC GAT ACT GAT GCA TAT GAG GGA GAT GTT TAT TTC     583
Gly Ile Ala Thr Ile Asp Thr Asp Ala Tyr Glu Gly Asp Val Tyr Phe
            360                 365                 370

AGG CTA AGG GTG AAT ATC ACA CTC TAT GAG TGT TAC TGATCCATGG          629
Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
        375                 380

GTATTCTATG TAGAATAGCT CAATCTGATA TGGCTATATT ATTTTGAGAG CATAGGTAGT   689

TAAGTTTTAT AACTAAGTAG TGAACCATGA GATCATTGAA AACTTGGGTG CTCATGCACA   749

GTTTTCATAT TTTCTAAATA AGTCTGCTCG ACTATTACAT TTATGGATTG TTGAGAATTG   809
```

```
TGTCGCTTAT TACTTTATGA ATAAGCTATT TTAAACAAAG TTTTCACAAG TTTAAAAGTT    869

GTCAAAAAAA AAAAAAAAAA A                                              890
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 192 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
Met Ala Ile Trp Asn Gly Arg Val Leu Asn Leu Cys Ile Leu Trp Leu
1               5                   10                  15

Leu Val Ser Ile Val Leu Leu Asn Gly Ile Asp Cys His Ser Arg Lys
            20                  25                  30

Lys Lys Leu Pro Lys Pro Cys Arg Asn Leu Val Leu Tyr Phe His Asp
        35                  40                  45

Ile Ile Tyr Asn Gly Lys Asn Ala Gly Asn Ala Thr Ser Thr Leu Val
    50                  55                  60

Ala Ala Pro Gln Gly Ala Asn Leu Thr Ile Met Thr Gly Asn Tyr His
65                  70                  75                  80

Phe Gly Asp Leu Ala Val Phe Asp Asp Pro Ile Thr Val Asp Asn Asn
                85                  90                  95

Leu His Ser Pro Pro Val Gly Arg Ala Gln Gly Phe Tyr Phe Tyr Asp
            100                 105                 110

Met Lys Asn Thr Phe Ser Ala Trp Leu Gly Phe Thr Phe Val Leu Asn
        115                 120                 125

Ser Thr Asp Tyr Lys Gly Thr Ile Thr Phe Gly Gly Ala Asp Pro Ile
    130                 135                 140

Leu Ala Lys Tyr Arg Asp Ile Ser Val Val Gly Gly Thr Gly Asp Phe
145                 150                 155                 160

Leu Met Ala Arg Gly Ile Ala Thr Ile Asp Thr Asp Ala Tyr Glu Gly
                165                 170                 175

Asp Val Tyr Phe Arg Leu Arg Val Asn Ile Thr Leu Tyr Glu Cys Tyr
            180                 185                 190
```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: N-terminal sequence from Forsythia intermedia (+)-pinoresinol/(+)-lariciresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly Arg
1               5                   10                  15

Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr
            20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal tryptic fragment from Forsythia intermedia (+)-pinoresinol/(+)-lariciresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```
Phe Met Asp Ile Ala Met Xaa Pro Gly Lys Val Thr Leu Asp Glu Lys
1               5                   10                  15
```

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal tryptic fragment from Forsythia intermedia (+)-pinoresinol/(+)-lariciresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Leu Pro Xaa Glu Phe Gly Met Asp Pro Ala Lys Phe Met
1 5 10

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal tryptic fragment from Forsythia intermedia (+)-pinoresinol/(+)-lariciresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Glu Val Val Gln Xaa Xaa Glu Lys
1 5

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal tryptic fragment from Forsythia intermedia (+)-pinoresinol/(+)-lariciresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Tyr Xaa Ser Val Glu Glu Tyr Leu Lys Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal cyanogen bromide fragment from Forsythia intermedia (+)-pinoresinol/(+)-lariciresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met
1 5 10

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal cyanogen bromide fragment from Forsythia intermedia (+)-pinoresinol/(+)-lariciresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Met Asp Pro Ala Lys Phe Met
1 5

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal cyanogen bromide fragment from Forsythia intermedia (+)-pinoresinol/(+)-lariciresinol reductase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```
Met Leu Ile Ser Phe Lys Met
1               5
```

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION:"PCR primer PLRN5"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

```
ATHATHGGNG GNACNGGNTA                                                20
```

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION:"PCR primer PLR14R"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```
GYTCCATNGC NATRTCCAT                                                 19
```

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION:"PCR primer PLR15R"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```
TCYTCNARNG TNACYTTNCC                                                20
```


(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1060 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia cDNA PLR-Fi1

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 28..963

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

```
AATTCGGCAC GAGAAAAACA GAGAGAG ATG GGA AAA AGC AAA GTT TTG ATC        51
                              Met Gly Lys Ser Lys Val Leu Ile
                                      195                 200

ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA AGT TTA      99
Ile Gly Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Lys Ala Ser Leu
            205                 210                 215

GCT CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GTT     147
Ala Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val
        220                 225                 230

GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT     195
Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys Met Gln Gly Ala
    235                 240                 245

CAT CTT GTA TCT GGT TCT TTC AAG GAT TTC AAC AGT CTG GTC GAG GCT     243
His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser Leu Val Glu Ala
250                 255                 260

GTC AAG CTC GTA GAC GTA GTA ATC AGC GCC ATT TCT GGT GTT CAT ATT     291
Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser Gly Val His Ile
265                 270                 275                 280

CGA AGC CAT CAA ATT CTT CTT CAA CTC AAG CTT GTT GAA GCT ATT AAA     339
Arg Ser His Gln Ile Leu Leu Gln Leu Lys Leu Val Glu Ala Ile Lys
                285                 290                 295

GAG GCT GGA AAT GTC AAG AGA TTT TTA CCA TCT GAG TTT GGA ATG GAT     387
Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp
            300                 305                 310

CCT GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA AAG GTA ACA CTT     435
Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu
        315                 320                 325

GAT GAG AAG ATG GTG GTA AGG AAA GCA ATT GAA AAG GCT GGG ATT CCT     483
Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro
    330                 335                 340
```

```
TTC ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT TTC TTG GGA GGT      531
Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly Gly
345                 350                 355                 360

CTC TGT CAA TTT GGC AAA ATT CTT CCT TCT AGA GAT TTT GTC ATT ATA      579
Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile Ile
                365                 370                 375

CAT GGA GAT GGT AAC AAA AAA GCA ATA TAT AAC AAT GAA GAT GAT ATA      627
His Gly Asp Gly Asn Lys Lys Ala Ile Tyr Asn Asn Glu Asp Asp Ile
            380                 385                 390

GCA ACT TAT GCC ATC AAA ACA ATT AAT GAT CCA AGA ACC CTC AAC AAG      675
Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys
        395                 400                 405

ACA ATC TAC ATT AGT CCT CCA AAA AAC ATC CTT TCA CAA AGA GAA GTT      723
Thr Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser Gln Arg Glu Val
    410                 415                 420

GTT CAG ACA TGG GAG AAG CTT ATT GGG AAA GAA CTG CAG AAA ATT ACA      771
Val Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu Gln Lys Ile Thr
425                 430                 435                 440

CTC TCG AAG GAA GAT TTT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT      819
Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Glu Tyr Ala
                445                 450                 455

CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC      867
Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Gln Gly Cys
            460                 465                 470

CTT ACG AGT TTT GAG ATA GGA GAT GAA GAA GAG GCA TCT AAA CTT TAT      915
Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Glu Ala Ser Lys Leu Tyr
        475                 480                 485

CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC AAG CGT TAC GTG      963
Pro Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu Lys Arg Tyr Val
    490                 495                 500

TAGTTGAAAG CTTTCCATTA TTATTGTAAT AATATTTAAA TCAGTATGTA GTTTTAAATT   1023

TCGTTAAATA ATATGTGTTG AATTTTGCTT CCAAAAA                            1060
```

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

```
Met Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly
  1               5                  10                  15
```



```
Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
             20                  25                  30

Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
         35                  40                  45

Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
     50                  55                  60

Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
 65                  70                  75                  80

Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
                 85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
            100                 105                 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
        115                 120                 125

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
    130                 135                 140

Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
145                 150                 155                 160

Phe Ala Gly Tyr Phe Leu Gly Gly Leu Cys Gln Phe Gly Lys Ile Leu
                165                 170                 175

Pro Ser Arg Asp Phe Val Ile Ile His Gly Asp Gly Asn Lys Lys Ala
            180                 185                 190

Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
        195                 200                 205

Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
    210                 215                 220

Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile
225                 230                 235                 240

Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
                245                 250                 255

Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
            260                 265                 270

His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
        275                 280                 285

Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
    290                 295                 300

Glu Glu Tyr Leu Lys Arg Tyr Val
305                 310
```

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1112 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia cDNA PLR-Fi2

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 44..979

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```
AATTCGGCAC GAGCTCGTGC CGCACAGAGA AAAACAGAGA GAG ATG GGA AAA AGC         55
                                                Met Gly Lys Ser
                                                        315

AAA GTT TTG ATC ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT        103
Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val
            320                 325                 330

AAG GCA AGT TTA GCT CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT        151
Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro
    335                 340                 345

GAA ATT GGT GTT GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA        199
Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys
350                 355                 360

ATG CAA GGA GCT CAT CTT GTA TCT GGT TCT TTC AAG GAT TTC AAC AGT        247
Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser
365                 370                 375                 380

CTG GTC GAG GCT GTC AAG CTC GTA GAC GTA GTA ATC AGC GCC ATT TCT        295
Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser
                385                 390                 395

GGT GTT CAT ATT CGA AGC CAT CAA ATT CTT CTT CAA CTC AAG CTT GTT        343
Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln Leu Lys Leu Val
            400                 405                 410

GAA GCT ATT AAA GAG GCT GGA AAT GTC AAG AGA TTT TTA CCA TCT GAG        391
Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu
        415                 420                 425

TTT GGA ATG GAT CCT GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA        439
Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly
        430                 435                 440

AAG GTA ACA CTT GAT GAG AAG ATG GTG GTA AGG AAA GCA ATT GAA AAG        487
Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys
445                 450                 455                 460
```



```
GCT GGG ATT CCT TTC ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT        535
Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr
                465                 470                 475

TTC TTG GGA GGT CTC TGT CAA TTT GGC AAA ATT CTT CCT TCT AGA GAT        583
Phe Leu Gly Gly Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp
            480                 485                 490

TTT GTC ATT ATA CAT GGA GAT GGT AAC AAA AAA GCA ATA TAT AAC AAT        631
Phe Val Ile Ile His Gly Asp Gly Asn Lys Lys Ala Ile Tyr Asn Asn
        495                 500                 505

GAA GAT GAT ATA GCA ACT TAT GCC ATC AAA ACA ATT AAT GAT CCA AGA        679
Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg
    510                 515                 520

ACC CTC AAC AAG ACA ATC TAC ATT AGT CCT CCA AAA AAC ATC CTT TCA        727
Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser
525                 530                 535                 540

CAA AGA GAA GTT GTT CAG ACA TGG GAG AAG CTT ATT GGG AAA GAA CTG        775
Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu
                545                 550                 555

CAG AAA ATT ACA CTC TCG AAG GAA GAT TTT TTA GCC TCC GTG AAA GAG        823
Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu
            560                 565                 570

CTC GAG TAT GCT CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC        871
Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn
        575                 580                 585

TAT CAG GGA TGC CTT ACG AGT TTT GAG ATA GGA GAT GAA GAA GAG GCA        919
Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Glu Ala
    590                 595                 600

TCT AAA CTT TAT CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC        967
Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu
605                 610                 615                 620

AAG CGT TAC GTG TAGTTGAAAG CTTTCCATTA TTATTGTAAT AATATTTAAA           1019
Lys Arg Tyr Val

TCAGTATGTA GTTTTAAATT TCGTTAAATA ATATGTGTTG AATTTTGCTT CAAACGAGTG      1079

GTCGATTGAA ATGGAATTTT GAAGTCAAAA AAA                                   1112
```

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

```
Met Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly
 1               5                  10                  15

Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
            20                  25                  30

Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
        35                  40                  45

Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
    50                  55                  60

Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
65                  70                  75                  80

Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
                85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
            100                 105                 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
        115                 120                 125

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
    130                 135                 140

Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
145                 150                 155                 160

Phe Ala Gly Tyr Phe Leu Gly Gly Leu Cys Gln Phe Gly Lys Ile Leu
                165                 170                 175

Pro Ser Arg Asp Phe Val Ile Ile His Gly Asp Gly Asn Lys Lys Ala
            180                 185                 190

Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
        195                 200                 205

Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
    210                 215                 220

Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile
225                 230                 235                 240

Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
                245                 250                 255

Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
            260                 265                 270

His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
        275                 280                 285

Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
    290                 295                 300

Glu Glu Tyr Leu Lys Arg Tyr Val
305                 310
```


(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1124 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia cDNA PLR-Fi3

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 29..964

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

```
AATTCGGCAC GAGGAAAAAC AGAGAGAG ATG GGA AAA AGC AAA GTT TTG ATC      52
                               Met Gly Lys Ser Lys Val Leu Ile
                                           315                 320

ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA AGT TTA    100
Ile Gly Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Lys Ala Ser Leu
                325                 330                 335

GCT CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GTT    148
Ala Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val
            340                 345                 350

GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT    196
Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys Met Gln Gly Ala
        355                 360                 365

CAT CTT GTA TCT GGT TCT TTC AAG GAT TTC AAC AGT CTG GTC GAG GCT    244
His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser Leu Val Glu Ala
    370                 375                 380

GTC AAG CTC GTA GAC GTA GTA ATC AGC GCC ATT TCT GGT GTT CAT ATT    292
Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser Gly Val His Ile
385                 390                 395                 400

CGA AGC CAT CAA ATT CTT CTT CAA CTC AAG CTT GTT GAA GCT ATT AAA    340
Arg Ser His Gln Ile Leu Leu Gln Leu Lys Leu Val Glu Ala Ile Lys
                405                 410                 415

GAG GCT GGA AAT GTC AAG AGA TTT TTA CCA TCT GAG TTT GGA ATG GAT    388
Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp
            420                 425                 430

CCT GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA AAG GTA ACA CTT    436
Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu
        435                 440                 445

GAT GAG AAG ATG GTG GTA AGG AAA GCA ATT GAA AAG GCT GGG ATT CCT    484
Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro
    450                 455                 460
```

```
TTC ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT TTC TTG GGA GGT      532
Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly Gly
465                 470                 475                 480

CTC TGT CAA TTT GGC AAA ATT CTT CCT TCT AGA GAT TTT GTC ATT ATA      580
Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile Ile
                485                 490                 495

CAT GGA GAT GGT AAC AAA AAA GCA ATA TAT AAC AAT GAA GAT GAT ATA      628
His Gly Asp Gly Asn Lys Lys Ala Ile Tyr Asn Asn Glu Asp Asp Ile
            500                 505                 510

GCA ACT TAT GCC ATC AAA ACA ATT AAT GAT CCA AGA ACC CTC AAC AAG      676
Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys
        515                 520                 525

ACA ATC TAC ATT AGT CCT CCA AAA AAC ATC CTT TCA CAA AGA GAA GTT      724
Thr Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser Gln Arg Glu Val
    530                 535                 540

GTT CAG ACA TGG GAG AAG CTT ATT GGG AAA GAA CTG CAG AAA ATT ACA      772
Val Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu Gln Lys Ile Thr
545                 550                 555                 560

CTC TCG AAG GAA GAT TTT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT      820
Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Glu Tyr Ala
                565                 570                 575

CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC      868
Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Gln Gly Cys
            580                 585                 590

CTT ACG AGT TTT GAG ATA GGA GAT GAA GAA GAG GCA TCT AAA CTT TAT      916
Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Glu Ala Ser Lys Leu Tyr
        595                 600                 605

CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC AAG CGT TAC GTG      964
Pro Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu Lys Arg Tyr Val
    610                 615                 620

TAGTTGAAAG CTTTCCATTA TTATTGTAAT AATATTTAAA TCAGTATGTA GTTTTAAATT   1024

TCGTTAAATA ATATGTGTTG AATTTTGCTT CAAACGAGTG GTCGATTGAA ATGGAATTTT   1084

GAAGTCATCT TCTCCACAAT ATTAGTCCAA ATAAAAAAAA                         1124
```

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:


```
Met Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly
1               5                   10                  15

Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
            20                  25                  30

Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
        35                  40                  45

Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
    50                  55                  60

Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
65                  70                  75                  80

Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
                85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
            100                 105                 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
        115                 120                 125

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
    130                 135                 140

Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
145                 150                 155                 160

Phe Ala Gly Tyr Phe Leu Gly Gly Leu Cys Gln Phe Gly Lys Ile Leu
                165                 170                 175

Pro Ser Arg Asp Phe Val Ile Ile His Gly Asp Gly Asn Lys Lys Ala
            180                 185                 190

Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
        195                 200                 205

Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
    210                 215                 220

Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile
225                 230                 235                 240

Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
                245                 250                 255

Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
            260                 265                 270

His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
        275                 280                 285

Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
    290                 295                 300

Glu Glu Tyr Leu Lys Arg Tyr Val
305                 310
```

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1097 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia cDNA PLR-Fi4

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 29..964

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

```
AATTCGGCAC GAGGAAAAAC AGAGAGAG ATG GGA AAA AGC AAA GTT TTG ATC      52
                               Met Gly Lys Ser Lys Val Leu Ile
                                        315                 320

ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA AGT TTA    100
Ile Gly Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Lys Ala Ser Leu
                325                 330                 335

GCT CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GTT    148
Ala Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val
            340                 345                 350

GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT    196
Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys Met Gln Gly Ala
        355                 360                 365

CAT CTT GTA TCT GGT TCT TTC AAG GAT TTC AAC AGT CTG GTC GAG GCT    244
His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser Leu Val Glu Ala
    370                 375                 380

GTC AAG CTC GTA GAC GTA GTA ATC AGC GCC ATT TCT GGT GTT CAT ATT    292
Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser Gly Val His Ile
385                 390                 395                 400

CGA AGC CAT CAA ATT CTT CTT CAA CTC AAG CTT GTT GAA GCT ATT AAA    340
Arg Ser His Gln Ile Leu Leu Gln Leu Lys Leu Val Glu Ala Ile Lys
                405                 410                 415

GAG GCT GGA AAT GTC AAG AGA TTT TTA CCA TCT GAG TTT GGA ATG GAT    388
Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp
            420                 425                 430

CCT GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA AAG GTA ACA CTT    436
Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu
        435                 440                 445

GAT GAG AAG ATG GTG GTA AGG AAA GCA ATT GAA AAG GCT GGG ATT CCT    484
Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro
    450                 455                 460
```



```
TTC ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT TTC TTG GGA GGT    532
Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly Gly
465                 470                 475                 480

CTC TGT CAA TTT GGC AAA ATT CTT CCT TCT AGA GAT TTT GTC ATT ATA    580
Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile Ile
                485                 490                 495

CAT GGA GAT GGT AAC AAA AAA GCA ATA TAT AAC AAT GAA GAT GAT ATA    628
His Gly Asp Gly Asn Lys Lys Ala Ile Tyr Asn Asn Glu Asp Asp Ile
            500                 505                 510

GCA ACT TAT GCC ATC AAA ACA ATT AAT GAT CCA AGA ACC CTC AAC AAG    676
Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys
        515                 520                 525

ACA ATC TAC ATT AGT CCT CCA AAA AAC ATC CTT TCA CAA AGA GAA GTT    724
Thr Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser Gln Arg Glu Val
    530                 535                 540

GTT CAG ACA TGG GAG AAG CTT ATT GGG AAA GAA CTG CAG AAA ATT ACA    772
Val Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu Gln Lys Ile Thr
545                 550                 555                 560

CTC TCG AAG GAA GAT TTT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT    820
Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Glu Tyr Ala
                565                 570                 575

CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC    868
Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Gln Gly Cys
            580                 585                 590

CTT ACG AGT TTT GAG ATA GGA GAT GAA GAA GAG GCA TCT AAA CTT TAT    916
Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Glu Ala Ser Lys Leu Tyr
        595                 600                 605

CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC AAG CGT TAC GTG    964
Pro Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu Lys Arg Tyr Val
    610                 615                 620

TAGTTGAAAG CTTTCCATTA TTATTGTAAT AATATTTAAA TCAGTATGTA GTTTTAAATT  1024

TCGTTAAATA ATATGTGTTG AATTTTGCTT CAAACGAGTG GTCGATTGAA ATGGAATTTT  1084

GAAAAAAAAA AAA                                                     1097
```

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

```
Met Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly
  1               5                  10                  15

Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
             20                  25                  30

Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
         35                  40                  45

Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
     50                  55                  60

Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
 65                  70                  75                  80

Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
                 85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
            100                 105                 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
        115                 120                 125

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
    130                 135                 140

Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
145                 150                 155                 160

Phe Ala Gly Tyr Phe Leu Gly Gly Leu Cys Gln Phe Gly Lys Ile Leu
                165                 170                 175

Pro Ser Arg Asp Phe Val Ile Ile His Gly Asp Gly Asn Lys Lys Ala
            180                 185                 190

Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
        195                 200                 205

Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
    210                 215                 220

Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile
225                 230                 235                 240

Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
                245                 250                 255

Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
            260                 265                 270

His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
        275                 280                 285

Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
    290                 295                 300

Glu Glu Tyr Leu Lys Arg Tyr Val
305                 310
```


(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1109 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia cDNA PLR-Fi5

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 31..966

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

```
AATTCGGCAC GAGGAGAAAA ACAGAGAGAG ATG GGA AAA AGC AAA GTT TTG ATC      54
                                 Met Gly Lys Ser Lys Val Leu Ile
                                         315                 320

ATT GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA AGT TTA      102
Ile Gly Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Lys Ala Ser Leu
                325                 330                 335

GCT CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GTT      150
Ala Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val
            340                 345                 350

GAT ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT      198
Asp Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys Met Gln Gly Ala
        355                 360                 365

CAT CTT GTA TCT GGT TCT TTC AAG GAT TTC AAC AGT CTG GTC GAG GCT      246
His Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser Leu Val Glu Ala
    370                 375                 380

GTC AAG CTC GTA GAC GTA GTA ATC AGC GCC ATT TCT GGT GTT CAT ATT      294
Val Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser Gly Val His Ile
385                 390                 395                 400

CGA AGC CAT CAA ATT CTT CTT CAA CTC AAG CTT GTT GAA GCT ATT AAA      342
Arg Ser His Gln Ile Leu Leu Gln Leu Lys Leu Val Glu Ala Ile Lys
                405                 410                 415

GAG GCT GGA AAT GTC AAG AGA TTT TTA CCA TCT GAG TTT GGA ATG GAT      390
Glu Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp
            420                 425                 430

CCT GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA AAG GTA ACA CTT      438
Pro Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu
        435                 440                 445

GAT GAG AAG ATG GTG GTA AGG AAA GCA ATT GAA AAG GCT GGG ATT CCT      486
Asp Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro
    450                 455                 460
```

```
TTC ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT TTC TTG GGA GGT      534
Phe Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly Gly
465                 470                 475                 480

CTC TGT CAA TTT GGC AAA ATT CTT CCT TCT AGA GAT TTT GTC ATT ATA      582
Leu Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile Ile
                485                 490                 495

CAT GGA GAT GGT AAC AAA AAA GCA ATA TAT AAC AAT GAA GAT GAT ATA      630
His Gly Asp Gly Asn Lys Lys Ala Ile Tyr Asn Asn Glu Asp Asp Ile
            500                 505                 510

GCA ACT TAT GCC ATC AAA ACA ATT AAT GAT CCA AGA ACC CTC AAC AAG      678
Ala Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys
        515                 520                 525

ACA ATC TAC ATT AGT CCT CCA AAA AAC ATC CTT TCA CAA AGA GAA GTT      726
Thr Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser Gln Arg Glu Val
    530                 535                 540

GTT CAG ACA TGG GAG AAG CTT ATT GGG AAA GAA CTG CAG AAA ATT ACA      774
Val Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu Gln Lys Ile Thr
545                 550                 555                 560

CTC TCG AAG GAA GAT TTT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT      822
Leu Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Glu Tyr Ala
                565                 570                 575

CAG CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC      870
Gln Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Gln Gly Cys
            580                 585                 590

CTT ACG AGT TTT GAG ATA GGA GAT GAA GAA GAG GCA TCT AAA CTT TAT      918
Leu Thr Ser Phe Glu Ile Gly Asp Glu Glu Glu Ala Ser Lys Leu Tyr
        595                 600                 605

CCA GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC AAG CGT TAC GTG      966
Pro Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu Lys Arg Tyr Val
    610                 615                 620

TAGTTGAAAG CTTTCCATTA TTATTGTAAT AATATTTAAA TCAGTATGTA GTTTTAAATT   1026

TCGTTAAATA ATATGTGTTG AATTTTGCTT CAAACGAGTG GTCGATTGAA ATGGAATTTT   1086

GAAGTCATCT TCTCCAAAAA AAA                                           1109
```

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:


```
Met Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly
1               5                   10                  15

Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
            20                  25                  30

Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
        35                  40                  45

Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
    50                  55                  60

Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
65                  70                  75                  80

Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
                85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
            100                 105                 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
        115                 120                 125

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
    130                 135                 140

Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
145                 150                 155                 160

Phe Ala Gly Tyr Phe Leu Gly Gly Leu Cys Gln Phe Gly Lys Ile Leu
                165                 170                 175

Pro Ser Arg Asp Phe Val Ile Ile His Gly Asp Gly Asn Lys Lys Ala
            180                 185                 190

Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
        195                 200                 205

Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
    210                 215                 220

Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile
225                 230                 235                 240

Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
                245                 250                 255

Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
            260                 265                 270

His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
        275                 280                 285

Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
    290                 295                 300

Glu Glu Tyr Leu Lys Arg Tyr Val
305                 310
```

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1107 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Forsythia intermedia cDNA PLR-Fi6

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 27..962

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

```
AATTCGGCAC GAGAAAACAG AGAGAG ATG GGA AAA AGC AAA GTT TTG ATC ATT      53
                             Met Gly Lys Ser Lys Val Leu Ile Ile
                                         315                 320

GGG GGT ACA GGG TAC TTA GGG AGG AGA TTG GTT AAG GCA AGT TTA GCT      101
Gly Gly Thr Gly Tyr Leu Gly Arg Arg Leu Val Lys Ala Ser Leu Ala
        325                 330                 335

CAA GGT CAT GAA ACA TAC ATT CTG CAT AGG CCT GAA ATT GGT GTT GAT      149
Gln Gly His Glu Thr Tyr Ile Leu His Arg Pro Glu Ile Gly Val Asp
    340                 345                 350

ATT GAT AAA GTT GAA ATG CTA ATA TCA TTT AAA ATG CAA GGA GCT CAT      197
Ile Asp Lys Val Glu Met Leu Ile Ser Phe Lys Met Gln Gly Ala His
355                 360                 365

CTT GTA TCT GGT TCT TTC AAG GAT TTC AAC AGT CTG GTC GAG GCT GTC      245
Leu Val Ser Gly Ser Phe Lys Asp Phe Asn Ser Leu Val Glu Ala Val
370                 375                 380                 385

AAG CTC GTA GAC GTA GTA ATC AGC GCC ATT TCT GGT GTT CAT ATT CGA      293
Lys Leu Val Asp Val Val Ile Ser Ala Ile Ser Gly Val His Ile Arg
                390                 395                 400

AGC CAT CAA ATT CTT CTT CAA CTC AAG CTT GTT GAA GCT ATT AAA GAG      341
Ser His Gln Ile Leu Leu Gln Leu Lys Leu Val Glu Ala Ile Lys Glu
            405                 410                 415

GCT GGA AAT GTC AAG AGA TTT TTA CCA TCT GAG TTT GGA ATG GAT CCT      389
Ala Gly Asn Val Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp Pro
        420                 425                 430

GCA AAA TTT ATG GAT ACG GCC ATG GAA CCC GGA AAG GTA ACA CTT GAT      437
Ala Lys Phe Met Asp Thr Ala Met Glu Pro Gly Lys Val Thr Leu Asp
    435                 440                 445

GAG AAG ATG GTG GTA AGG AAA GCA ATT GAA AAG GCT GGG ATT CCT TTC      485
Glu Lys Met Val Val Arg Lys Ala Ile Glu Lys Ala Gly Ile Pro Phe
450                 455                 460                 465
```



```
ACA TAT GTC TCT GCA AAT TGC TTT GCT GGT TAT TTC TTG GGA GGT CTC      533
Thr Tyr Val Ser Ala Asn Cys Phe Ala Gly Tyr Phe Leu Gly Gly Leu
                470                 475                 480

TGT CAA TTT GGC AAA ATT CTT CCT TCT AGA GAT TTT GTC ATT ATA CAT      581
Cys Gln Phe Gly Lys Ile Leu Pro Ser Arg Asp Phe Val Ile Ile His
            485                 490                 495

GGA GAT GGT AAC AAA AAA GCA ATA TAT AAC AAT GAA GAT GAT ATA GCA      629
Gly Asp Gly Asn Lys Lys Ala Ile Tyr Asn Asn Glu Asp Asp Ile Ala
        500                 505                 510

ACT TAT GCC ATC AAA ACA ATT AAT GAT CCA AGA ACC CTC AAC AAG ACA      677
Thr Tyr Ala Ile Lys Thr Ile Asn Asp Pro Arg Thr Leu Asn Lys Thr
    515                 520                 525

ATC TAC ATT AGT CCT CCA AAA AAC ATC CTT TCA CAA AGA GAA GTT GTT      725
Ile Tyr Ile Ser Pro Pro Lys Asn Ile Leu Ser Gln Arg Glu Val Val
530                 535                 540                 545

CAG ACA TGG GAG AAG CTT ATT GGG AAA GAA CTG CAG AAA ATT ACA CTC      773
Gln Thr Trp Glu Lys Leu Ile Gly Lys Glu Leu Gln Lys Ile Thr Leu
                550                 555                 560

TCG AAG GAA GAT TTT TTA GCC TCC GTG AAA GAG CTC GAG TAT GCT CAG      821
Ser Lys Glu Asp Phe Leu Ala Ser Val Lys Glu Leu Glu Tyr Ala Gln
            565                 570                 575

CAA GTG GGA TTA AGC CAT TAT CAT GAT GTC AAC TAT CAG GGA TGC CTT      869
Gln Val Gly Leu Ser His Tyr His Asp Val Asn Tyr Gln Gly Cys Leu
        580                 585                 590

ACG AGT TTT GAG ATA GGA GAT GAA GAA GAG GCA TCT AAA CTT TAT CCA      917
Thr Ser Phe Glu Ile Gly Asp Glu Glu Glu Ala Ser Lys Leu Tyr Pro
    595                 600                 605

GAG GTT AAG TAT ACC AGT GTG GAA GAG TAC CTC AAG CGT TAC GTG          962
Glu Val Lys Tyr Thr Ser Val Glu Glu Tyr Leu Lys Arg Tyr Val
610                 615                 620

TAGTTGAAAG CTTTCCATTA TTATTGTAAT AATATTTAAA TCAGTATGTA GTTTTAAATT    1022

TCGTTAAATA ATATGTGTTG AATTTTGCTT CAAACGAGTG GTCGATTGAA ATGGAATTTT    1082

GAAGTCATCT TCTCCACAAA AAAAA                                          1107
```

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

```
Met Gly Lys Ser Lys Val Leu Ile Ile Gly Gly Thr Gly Tyr Leu Gly
  1               5                  10                  15

Arg Arg Leu Val Lys Ala Ser Leu Ala Gln Gly His Glu Thr Tyr Ile
            20                  25                  30

Leu His Arg Pro Glu Ile Gly Val Asp Ile Asp Lys Val Glu Met Leu
        35                  40                  45

Ile Ser Phe Lys Met Gln Gly Ala His Leu Val Ser Gly Ser Phe Lys
    50                  55                  60

Asp Phe Asn Ser Leu Val Glu Ala Val Lys Leu Val Asp Val Val Ile
65                  70                  75                  80

Ser Ala Ile Ser Gly Val His Ile Arg Ser His Gln Ile Leu Leu Gln
                85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Val Lys Arg Phe
            100                 105                 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Ala Lys Phe Met Asp Thr Ala
        115                 120                 125

Met Glu Pro Gly Lys Val Thr Leu Asp Glu Lys Met Val Val Arg Lys
    130                 135                 140

Ala Ile Glu Lys Ala Gly Ile Pro Phe Thr Tyr Val Ser Ala Asn Cys
145                 150                 155                 160

Phe Ala Gly Tyr Phe Leu Gly Gly Leu Cys Gln Phe Gly Lys Ile Leu
                165                 170                 175

Pro Ser Arg Asp Phe Val Ile Ile His Gly Asp Gly Asn Lys Lys Ala
            180                 185                 190

Ile Tyr Asn Asn Glu Asp Asp Ile Ala Thr Tyr Ala Ile Lys Thr Ile
        195                 200                 205

Asn Asp Pro Arg Thr Leu Asn Lys Thr Ile Tyr Ile Ser Pro Pro Lys
    210                 215                 220

Asn Ile Leu Ser Gln Arg Glu Val Val Gln Thr Trp Glu Lys Leu Ile
225                 230                 235                 240

Gly Lys Glu Leu Gln Lys Ile Thr Leu Ser Lys Glu Asp Phe Leu Ala
                245                 250                 255

Ser Val Lys Glu Leu Glu Tyr Ala Gln Gln Val Gly Leu Ser His Tyr
            260                 265                 270

His Asp Val Asn Tyr Gln Gly Cys Leu Thr Ser Phe Glu Ile Gly Asp
        275                 280                 285

Glu Glu Glu Ala Ser Lys Leu Tyr Pro Glu Val Lys Tyr Thr Ser Val
    290                 295                 300

Glu Glu Tyr Leu Lys Arg Tyr Val
305                 310
```


(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION:"cDNA synthesis linker primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

```
GTCTCGAGTT TTTTTTTTT TTTTTT                                  26
```

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: "cDNA synthesis primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

```
GCACATAAGA GTATGGATAA G                                      21
```

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1190 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata cDNA PLR-Tp1

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 13..951

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

```
GCACATAAGA GT ATG GAT AAG AAG AGC AGA GTT CTG ATA GTG GGG GGC        48
              Met Asp Lys Lys Ser Arg Val Leu Ile Val Gly Gly
                      315                 320
```

```
ACT GGT TAT ATA GGC AAA AGA ATT GTG AAT GCC AGT ATA TCT CTT GGC       96
Thr Gly Tyr Ile Gly Lys Arg Ile Val Asn Ala Ser Ile Ser Leu Gly
325                 330                 335                 340

CAT CCC ACT TAT GTT TTG TTC AGA CCA GAA GTG GTC TCT AAC ATT GAC      144
His Pro Thr Tyr Val Leu Phe Arg Pro Glu Val Val Ser Asn Ile Asp
                345                 350                 355

AAA GTG CAG ATG CTG TTA TAC TTC AAA CAG CTT GGT GCC AAA CTT ATT      192
Lys Val Gln Met Leu Leu Tyr Phe Lys Gln Leu Gly Ala Lys Leu Ile
            360                 365                 370

GAG GCT TCA TTG GAT GAC CAC CAA AGG CTT GTG GAT GCT CTG AAA CAA      240
Glu Ala Ser Leu Asp Asp His Gln Arg Leu Val Asp Ala Leu Lys Gln
        375                 380                 385

GTG GAT GTT GTC ATA AGT GCT TTG GCA GGA GGT GTT CTA AGC CAC CAT      288
Val Asp Val Val Ile Ser Ala Leu Ala Gly Gly Val Leu Ser His His
    390                 395                 400

ATA CTT GAA CAG CTC AAA CTA GTG GAA GCC ATC AAA GAA GCT GGA AAT      336
Ile Leu Glu Gln Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn
405                 410                 415                 420

ATT AAG AGA TTT CTT CCA TCT GAG TTT GGC ATG GAT CCA GAT ATT ATG      384
Ile Lys Arg Phe Leu Pro Ser Glu Phe Gly Met Asp Pro Asp Ile Met
                425                 430                 435

GAG CAT GCA TTG CAA CCT GGT AGC ATT ACA TTC ATC GAT AAG AGA AAG      432
Glu His Ala Leu Gln Pro Gly Ser Ile Thr Phe Ile Asp Lys Arg Lys
            440                 445                 450

GTT CGG CGT GCC ATT GAA GCA GCA TCC ATT CCT TAC ACA TAT GTG TCT      480
Val Arg Arg Ala Ile Glu Ala Ala Ser Ile Pro Tyr Thr Tyr Val Ser
        455                 460                 465

TCA AAT ATG TTT GCT GGT TAC TTT GCT GGA AGT TTA GCT CAA CTT GAT      528
Ser Asn Met Phe Ala Gly Tyr Phe Ala Gly Ser Leu Ala Gln Leu Asp
    470                 475                 480

GGT CAT ATG ATG CCT CCT CGA GAC AAG GTC CTC ATC TAT GGA GAT GGA      576
Gly His Met Met Pro Pro Arg Asp Lys Val Leu Ile Tyr Gly Asp Gly
485                 490                 495                 500

AAT GTT AAA GGT ATT TGG GTG GAT GAA GAT GAT GTT GGA ACA TAC ACA      624
Asn Val Lys Gly Ile Trp Val Asp Glu Asp Asp Val Gly Thr Tyr Thr
                505                 510                 515

ATC AAA TCA ATT GAT GAT CCA CAA ACC CTT AAC AAG ACT ATG TAT ATT      672
Ile Lys Ser Ile Asp Asp Pro Gln Thr Leu Asn Lys Thr Met Tyr Ile
            520                 525                 530

AGG CCA CCT ATG AAT ATC CTT TCA CAG AAG GAA GTT ATA CAA ATA TGG      720
Arg Pro Pro Met Asn Ile Leu Ser Gln Lys Glu Val Ile Gln Ile Trp
        535                 540                 545

GAG AGA TTA TCA GAA CAA AAC CTG GAT AAA ATA TAC ATT TCT TCT CAA      768
Glu Arg Leu Ser Glu Gln Asn Leu Asp Lys Ile Tyr Ile Ser Ser Gln
    550                 555                 560
```



```
GAC TTT CTT GCA GAT ATG AAA GAT AAA TCA TAT GAA GAG AAG ATT GTA      816
Asp Phe Leu Ala Asp Met Lys Asp Lys Ser Tyr Glu Glu Lys Ile Val
565                 570                 575                 580

CGA TGT CAT CTC TAC CAA ATT TTC TTT AGA GGA GAT CTT TAC AAC TTT      864
Arg Cys His Leu Tyr Gln Ile Phe Phe Arg Gly Asp Leu Tyr Asn Phe
                585                 590                 595

GAA ATT GGC CCC AAT GCT ATT GAA GCT ACC AAA CTT TAT CCA GAA GTG      912
Glu Ile Gly Pro Asn Ala Ile Glu Ala Thr Lys Leu Tyr Pro Glu Val
            600                 605                 610

AAA TAC GTA ACC ATG GAT TCA TAT TTA GAG CGC TAT GTT TGAATATCTT       961
Lys Tyr Val Thr Met Asp Ser Tyr Leu Glu Arg Tyr Val
        615                 620                 625

TCTAGTTTTG TATATTGTTT TTCTACATGA TAATGTGAGA GGTACTATTT CAAATAATTT   1021

AGACTTATGG CTCAATTTTA AAACTAGAGT ACACTTTATT CCAAATTACT TACACTATTT   1081

TTTACTTCAT ATTGTACTCA ATATAGACTT GGTATAAAGA ATATGGAATC ATAATGATAT   1141

TATAATTATT TATAGATCTT ATTTTAAATA AAAAAAAAAA AAAAAAAA                1190
```

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 313 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

```
Met Asp Lys Lys Ser Arg Val Leu Ile Val Gly Gly Thr Gly Tyr Ile
 1               5                  10                  15

Gly Lys Arg Ile Val Asn Ala Ser Ile Ser Leu Gly His Pro Thr Tyr
            20                  25                  30

Val Leu Phe Arg Pro Glu Val Val Ser Asn Ile Asp Lys Val Gln Met
        35                  40                  45

Leu Leu Tyr Phe Lys Gln Leu Gly Ala Lys Leu Ile Glu Ala Ser Leu
    50                  55                  60

Asp Asp His Gln Arg Leu Val Asp Ala Leu Lys Gln Val Asp Val Val
65                  70                  75                  80

Ile Ser Ala Leu Ala Gly Gly Val Leu Ser His His Ile Leu Glu Gln
                85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe
            100                 105                 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Asp Ile Met Glu His Ala Leu
        115                 120                 125
```

```
Gln Pro Gly Ser Ile Thr Phe Ile Asp Lys Arg Lys Val Arg Arg Ala
    130                 135                 140

Ile Glu Ala Ala Ser Ile Pro Tyr Thr Tyr Val Ser Ser Asn Met Phe
145                 150                 155                 160

Ala Gly Tyr Phe Ala Gly Ser Leu Ala Gln Leu Asp Gly His Met Met
                165                 170                 175

Pro Pro Arg Asp Lys Val Leu Ile Tyr Gly Asp Gly Asn Val Lys Gly
            180                 185                 190

Ile Trp Val Asp Glu Asp Asp Val Gly Thr Tyr Thr Ile Lys Ser Ile
        195                 200                 205

Asp Asp Pro Gln Thr Leu Asn Lys Thr Met Tyr Ile Arg Pro Pro Met
    210                 215                 220

Asn Ile Leu Ser Gln Lys Glu Val Ile Gln Ile Trp Glu Arg Leu Ser
225                 230                 235                 240

Glu Gln Asn Leu Asp Lys Ile Tyr Ile Ser Ser Gln Asp Phe Leu Ala
                245                 250                 255

Asp Met Lys Asp Lys Ser Tyr Glu Glu Lys Ile Val Arg Cys His Leu
            260                 265                 270

Tyr Gln Ile Phe Phe Arg Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro
        275                 280                 285

Asn Ala Ile Glu Ala Thr Lys Leu Tyr Pro Glu Val Lys Tyr Val Thr
    290                 295                 300

Met Asp Ser Tyr Leu Glu Arg Tyr Val
305                 310
```

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1151 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata cDNA PLR-Tp2

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 61..996

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

```
GATAAGCAGC ATTTCTTCAC CAAAGTGGTC CGCCATTAAA GGAATAGTTT GAAAGCAGAG      60
```



```
ATG GAA GAG AGT AGC AGG GTT TTG ATA GTG GGA GGC ACA GGA TAC ATA      108
Met Glu Glu Ser Ser Arg Val Leu Ile Val Gly Gly Thr Gly Tyr Ile
    315                 320                 325

GGC AGA AGG ATT GTG AAA GCC AGC ATT GCT CTG GGC CAT CCT ACT TTC      156
Gly Arg Arg Ile Val Lys Ala Ser Ile Ala Leu Gly His Pro Thr Phe
330                 335                 340                 345

ATT TTG TTT AGG AAA GAA GTT GTT TCT GAT GTA GAG AAA GTG GAG ATG      204
Ile Leu Phe Arg Lys Glu Val Val Ser Asp Val Glu Lys Val Glu Met
                350                 355                 360

TTA TTG TCC TTC AAA AAG AAT GGT GCC AAA TTA CTG GAG GCT TCA TTT      252
Leu Leu Ser Phe Lys Lys Asn Gly Ala Lys Leu Leu Glu Ala Ser Phe
            365                 370                 375

GAT GAT CAC GAA AGC CTT GTA GAT GCT GTG AAG CAG GTT GAT GTT GTG      300
Asp Asp His Glu Ser Leu Val Asp Ala Val Lys Gln Val Asp Val Val
        380                 385                 390

ATA AGT GCA GTT GCA GGA AAC CAC ATG CGG CAT CAC ATC CTT CAA CAG      348
Ile Ser Ala Val Ala Gly Asn His Met Arg His His Ile Leu Gln Gln
    395                 400                 405

CTC AAA TTA GTG GAG GCC ATT AAA GAA GCT GGA AAT ATT AAG AGG TTT      396
Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe
410                 415                 420                 425

GTT CCT TCA GAA TTT GGG ATG GAT CCA GGG TTA ATG GAG CAT GCA ATG      444
Val Pro Ser Glu Phe Gly Met Asp Pro Gly Leu Met Glu His Ala Met
                430                 435                 440

GCA CCT GGC AAC ATT GTA TTT ATT GAT AAA ATA AAA GTT CGA GAG GCC      492
Ala Pro Gly Asn Ile Val Phe Ile Asp Lys Ile Lys Val Arg Glu Ala
            445                 450                 455

ATA GAA GCT GCA TCC ATT CCT CAC ACT TAT ATC TCT GCC AAC ATA TTT      540
Ile Glu Ala Ala Ser Ile Pro His Thr Tyr Ile Ser Ala Asn Ile Phe
        460                 465                 470

GCT GGC TAC TTG GTT GGT GGA TTA GCT CAA CTT GGT CGT GTG ATG CCT      588
Ala Gly Tyr Leu Val Gly Gly Leu Ala Gln Leu Gly Arg Val Met Pro
    475                 480                 485

CCT TCA GAA AAA GTA ATT CTC TAT GGA GAT GGA AAT GTC AAA GCT GTT      636
Pro Ser Glu Lys Val Ile Leu Tyr Gly Asp Gly Asn Val Lys Ala Val
490                 495                 500                 505

TGG GTA GAT GAA GAT GAT GTT GGA ATA TAC ACA ATC AAA GCA ATT GAT      684
Trp Val Asp Glu Asp Asp Val Gly Ile Tyr Thr Ile Lys Ala Ile Asp
                510                 515                 520

GAC CCT CAC ACC CTA AAT AAG ACT ATG TAC ATC AGG CCA CCT TTG AAT      732
Asp Pro His Thr Leu Asn Lys Thr Met Tyr Ile Arg Pro Pro Leu Asn
            525                 530                 535

ATT CTT TCT CAG AAG GAA GTG GTT GAA AAA TGG GAA AAG TTA TCA GGA      780
Ile Leu Ser Gln Lys Glu Val Val Glu Lys Trp Glu Lys Leu Ser Gly
        540                 545                 550
```

```
AAG AGC TTA AAT AAA ATA AAT ATT TCT GTT GAA GAT TTT CTT GCA GGC        828
Lys Ser Leu Asn Lys Ile Asn Ile Ser Val Glu Asp Phe Leu Ala Gly
    555                 560                 565

ATG GAA GGT CAA TCA TAT GGA GAG CAG ATT GGA ATA TCA CAT TTC TAC        876
Met Glu Gly Gln Ser Tyr Gly Glu Gln Ile Gly Ile Ser His Phe Tyr
570                 575                 580                 585

CAA ATG TTC TAT AGG GGT GAT CTT TAT AAT TTT GAA ATT GGA CCT AAT        924
Gln Met Phe Tyr Arg Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asn
                590                 595                 600

GGA GTA GAA GCT TCC CAA CTT TAT CCA GAA GTA AAA TAT ACA ACA GTG        972
Gly Val Glu Ala Ser Gln Leu Tyr Pro Glu Val Lys Tyr Thr Thr Val
            605                 610                 615

GAT TCA TAC ATG GAA CGC TAC CTA TGAAAATCTT CTTCACGAAG ATATCTAAAT      1026
Asp Ser Tyr Met Glu Arg Tyr Leu
        620                 625

TTAATTTAAG CTTTCTAAAA GTTTTTATAT TTTGACATTA TGCTAAATAA AAATGGAGAG      1086

TATCTAGATA ATAATATTGA CCAATCATAT TAAAAATTAT TGGGATTAAA AAAAAAAAAA      1146

AAAAA                                                                  1151
```

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

```
Met Glu Glu Ser Ser Arg Val Leu Ile Val Gly Gly Thr Gly Tyr Ile
  1               5                  10                  15

Gly Arg Arg Ile Val Lys Ala Ser Ile Ala Leu Gly His Pro Thr Phe
             20                  25                  30

Ile Leu Phe Arg Lys Glu Val Val Ser Asp Val Glu Lys Val Glu Met
         35                  40                  45

Leu Leu Ser Phe Lys Lys Asn Gly Ala Lys Leu Leu Glu Ala Ser Phe
    50                  55                  60

Asp Asp His Glu Ser Leu Val Asp Ala Val Lys Gln Val Asp Val Val
65                  70                  75                  80

Ile Ser Ala Val Ala Gly Asn His Met Arg His His Ile Leu Gln Gln
                85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe
            100                 105                 110
```



```
Val Pro Ser Glu Phe Gly Met Asp Pro Gly Leu Met Glu His Ala Met
        115                 120                 125

Ala Pro Gly Asn Ile Val Phe Ile Asp Lys Ile Lys Val Arg Glu Ala
    130                 135                 140

Ile Glu Ala Ala Ser Ile Pro His Thr Tyr Ile Ser Ala Asn Ile Phe
145                 150                 155                 160

Ala Gly Tyr Leu Val Gly Gly Leu Ala Gln Leu Gly Arg Val Met Pro
                165                 170                 175

Pro Ser Glu Lys Val Ile Leu Tyr Gly Asp Gly Asn Val Lys Ala Val
            180                 185                 190

Trp Val Asp Glu Asp Asp Val Gly Ile Tyr Thr Ile Lys Ala Ile Asp
        195                 200                 205

Asp Pro His Thr Leu Asn Lys Thr Met Tyr Ile Arg Pro Pro Leu Asn
    210                 215                 220

Ile Leu Ser Gln Lys Glu Val Val Glu Lys Trp Glu Lys Leu Ser Gly
225                 230                 235                 240

Lys Ser Leu Asn Lys Ile Asn Ile Ser Val Glu Asp Phe Leu Ala Gly
                245                 250                 255

Met Glu Gly Gln Ser Tyr Gly Glu Gln Ile Gly Ile Ser His Phe Tyr
            260                 265                 270

Gln Met Phe Tyr Arg Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asn
        275                 280                 285

Gly Val Glu Ala Ser Gln Leu Tyr Pro Glu Val Lys Tyr Thr Thr Val
    290                 295                 300

Asp Ser Tyr Met Glu Arg Tyr Leu
305                 310
```

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata cDNA PLR-Tp3

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 164..1105

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

```
AAAAACTCTT AGACTTATTT TCATTTTTAC CCAGTTCATA AGTGTTTGTT GGGTCTCTTC      60

AAAAAAAGCC CCCTCTCGTT AGAGGCAAAG AACAGCATGC TCAGATATAT GTAAGAAGCA     120

AAATGCCCAA AATTTGACTG TGAAAGTGGA TGCACATAAG AAT ATG GAT AAG AAG       175
                                                Met Asp Lys Lys
                                                            315

AGC AGA GTT CTA ATA GTG GGG GGT ACT GGT TTT ATA GGC AAA AGA ATT       223
Ser Arg Val Leu Ile Val Gly Gly Thr Gly Phe Ile Gly Lys Arg Ile
            320                 325                 330

GTG AAG GCC AGT TTG GCT CTT GGC CAT CCT ACT TAT GTT TTG TTC AGG       271
Val Lys Ala Ser Leu Ala Leu Gly His Pro Thr Tyr Val Leu Phe Arg
        335                 340                 345

CCA GAA GCC CTC TCT TAC ATT GAC AAA GTG CAG ATG TTG ATA TCC TTC       319
Pro Glu Ala Leu Ser Tyr Ile Asp Lys Val Gln Met Leu Ile Ser Phe
    350                 355                 360

AAA CAG CTT GGG GCC AAA CTT CTT GAG GCT TCA TTG GAT GAC CAC CAA       367
Lys Gln Leu Gly Ala Lys Leu Leu Glu Ala Ser Leu Asp Asp His Gln
365                 370                 375                 380

GGG CTT GTG GAT GTT GTG AAA CAA GTA GAT GTT GTG ATC AGT GCT GTT       415
Gly Leu Val Asp Val Val Lys Gln Val Asp Val Val Ile Ser Ala Val
                385                 390                 395

TCA GGA GGT CTG GTG CGC CAC CAT ATA CTT GAC CAG CTC AAG CTA GTG       463
Ser Gly Gly Leu Val Arg His His Ile Leu Asp Gln Leu Lys Leu Val
            400                 405                 410

GAG GCA ATT AAA GAA GCT GGC AAT ATT AAG AGA TTT CTT CCT TCA GAA       511
Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe Leu Pro Ser Glu
        415                 420                 425

TTT GGG ATG GAC CCA GAT GTT GTA GAA GAT CCA TTG GAA CCT GGT AAC       559
Phe Gly Met Asp Pro Asp Val Val Glu Asp Pro Leu Glu Pro Gly Asn
    430                 435                 440

ATT ACA TTC ATT GAT AAA AGA AAA GTT AGA CGT GCC ATT GAA GCA GCA       607
Ile Thr Phe Ile Asp Lys Arg Lys Val Arg Arg Ala Ile Glu Ala Ala
445                 450                 455                 460

ACC ATT CCT TAC ACA TAT GTG TCT TCA AAT ATG TTT GCT GGG TTC TTT       655
Thr Ile Pro Tyr Thr Tyr Val Ser Ser Asn Met Phe Ala Gly Phe Phe
                465                 470                 475

GCT GGA AGC TTA GCA CAA CTG CAA GAT GCT CCC CGC ATG ATG CCT GCT       703
Ala Gly Ser Leu Ala Gln Leu Gln Asp Ala Pro Arg Met Met Pro Ala
            480                 485                 490

CGA GAT AAA GTT CTC ATA TAT GGA GAT GGA AAT GTT AAA GGT GTT TAT       751
Arg Asp Lys Val Leu Ile Tyr Gly Asp Gly Asn Val Lys Gly Val Tyr
        495                 500                 505

GTA GAT GAA GAT GAT GCT GGA ATA TAC ATA GTC AAA TCA ATT GAT GAT       799
Val Asp Glu Asp Asp Ala Gly Ile Tyr Ile Val Lys Ser Ile Asp Asp
    510                 515                 520
```



```
CCT CGC ACA CTC AAC AAG ACT GTG TAT ATC AGG CCA CCA ATG AAT ATA       847
Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro Pro Met Asn Ile
525                 530                 535                 540

CTT TCA CAG AAA GAA GTA GTT GAA ATA TGG GAG AGA CTA TCA GGT TTG       895
Leu Ser Gln Lys Glu Val Val Glu Ile Trp Glu Arg Leu Ser Gly Leu
                545                 550                 555

AGC CTA GAA AAA ATC TAC GTT TCT GAG GAC CAA CTT CTT AAT ATG AAA       943
Ser Leu Glu Lys Ile Tyr Val Ser Glu Asp Gln Leu Leu Asn Met Lys
            560                 565                 570

GAT AAA TCT TAT GTG GAG AAG ATG GCA CGA TGT CAT CTC TAT CAT TTT       991
Asp Lys Ser Tyr Val Glu Lys Met Ala Arg Cys His Leu Tyr His Phe
        575                 580                 585

TTT ATC AAA GGG GAT CTT TAC AAT TTT GAA ATT GGA CCC AAT GCT ACT      1039
Phe Ile Lys Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asn Ala Thr
    590                 595                 600

GAA GGC ACA AAA CTT TAT CCA GAA GTC AAA TAC ACA ACC ATG GAT TCA      1087
Glu Gly Thr Lys Leu Tyr Pro Glu Val Lys Tyr Thr Thr Met Asp Ser
605                 610                 615                 620

TAT ATG GAG CGT TAT CTA TAGCTAATAG ATTTTTCTTA AATAATAGCT            1135
Tyr Met Glu Arg Tyr Leu
                625

TGAAATATTC TATACTCAAT AAGAGTGTAT TCATAAATAA TACACAACAC TTGCTCTTTT    1195

ATAGATTACT TTTTTAATAG GTGGCTTTTA TAAAACATGT ATAAAAAAAA TTGCAAACAA    1255

TATTTTTAAA TTAGCAATAA TAACCACCTT TAAATAAAAA AAAAAAAAAA AAA           1308
```

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 314 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```
Met Asp Lys Lys Ser Arg Val Leu Ile Val Gly Gly Thr Gly Phe Ile
1               5                   10                  15

Gly Lys Arg Ile Val Lys Ala Ser Leu Ala Leu Gly His Pro Thr Tyr
            20                  25                  30

Val Leu Phe Arg Pro Glu Ala Leu Ser Tyr Ile Asp Lys Val Gln Met
        35                  40                  45

Leu Ile Ser Phe Lys Gln Leu Gly Ala Lys Leu Leu Glu Ala Ser Leu
    50                  55                  60

Asp Asp His Gln Gly Leu Val Asp Val Val Lys Gln Val Asp Val Val
65                  70                  75                  80
```

```
Ile Ser Ala Val Ser Gly Gly Leu Val Arg His His Ile Leu Asp Gln
                85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe
            100                 105                 110

Leu Pro Ser Glu Phe Gly Met Asp Pro Asp Val Val Glu Asp Pro Leu
        115                 120                 125

Glu Pro Gly Asn Ile Thr Phe Ile Asp Lys Arg Lys Val Arg Arg Ala
    130                 135                 140

Ile Glu Ala Ala Thr Ile Pro Tyr Thr Tyr Val Ser Ser Asn Met Phe
145                 150                 155                 160

Ala Gly Phe Phe Ala Gly Ser Leu Ala Gln Leu Gln Asp Ala Pro Arg
                165                 170                 175

Met Met Pro Ala Arg Asp Lys Val Leu Ile Tyr Gly Asp Gly Asn Val
            180                 185                 190

Lys Gly Val Tyr Val Asp Glu Asp Asp Ala Gly Ile Tyr Ile Val Lys
        195                 200                 205

Ser Ile Asp Asp Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro
    210                 215                 220

Pro Met Asn Ile Leu Ser Gln Lys Glu Val Val Glu Ile Trp Glu Arg
225                 230                 235                 240

Leu Ser Gly Leu Ser Leu Glu Lys Ile Tyr Val Ser Glu Asp Gln Leu
                245                 250                 255

Leu Asn Met Lys Asp Lys Ser Tyr Val Glu Lys Met Ala Arg Cys His
            260                 265                 270

Leu Tyr His Phe Phe Ile Lys Gly Asp Leu Tyr Asn Phe Glu Ile Gly
        275                 280                 285

Pro Asn Ala Thr Glu Gly Thr Lys Leu Tyr Pro Glu Val Lys Tyr Thr
    290                 295                 300

Thr Met Asp Ser Tyr Met Glu Arg Tyr Leu
305                 310
```

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1287 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Thuja plicata cDNA PLR-Tp4

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 11..946

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```
GAAAGCAGAG ATG GAA GAG AGT AGC AGG ATT TTG GTA GTG GGA GGC ACA      49
           Met Glu Glu Ser Ser Arg Ile Leu Val Val Gly Gly Thr
           315                 320                 325

GGA TAC ATA GGC AGA AGG ATT GTG AAA GCC AGC ATT GCT CTG GGC CAT      97
Gly Tyr Ile Gly Arg Arg Ile Val Lys Ala Ser Ile Ala Leu Gly His
        330                 335                 340

CCT ACT TTC ATT TTG TTT AGG AAA GAA GTT GTT TCT GAT GTA GAG AAA     145
Pro Thr Phe Ile Leu Phe Arg Lys Glu Val Val Ser Asp Val Glu Lys
    345                 350                 355

GTG GAG ATG TTA TTG TCC TTC AAA AAG AAT GGT GCC AAA TTA CTG GAG     193
Val Glu Met Leu Leu Ser Phe Lys Lys Asn Gly Ala Lys Leu Leu Glu
360                 365                 370                 375

GCT TCA TTT GAT GAT CAC GAA AGC CTT GTA GAT GCT GTG AAG CAG GTT     241
Ala Ser Phe Asp Asp His Glu Ser Leu Val Asp Ala Val Lys Gln Val
                380                 385                 390

GAT GTT GTC ATA AGT GCA GTT GCA GGA AAC CAC ATG CGG CAT CAC ATC     289
Asp Val Val Ile Ser Ala Val Ala Gly Asn His Met Arg His His Ile
            395                 400                 405

CTT CAA CAG CTC AAA TTA GTG GAG GCC ATT AAA GAA GCT GGA AAT ATT     337
Leu Gln Gln Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile
        410                 415                 420

AAG AGG TTT GTC CCT TCA GAA TTT GGG ATG GAT CCA GGG TTA ATG GAC     385
Lys Arg Phe Val Pro Ser Glu Phe Gly Met Asp Pro Gly Leu Met Asp
    425                 430                 435

CAT GCA ATG GCA CCA GGA AAC ATT GTA TTT ATT GAT AAA ATA AAA GTT     433
His Ala Met Ala Pro Gly Asn Ile Val Phe Ile Asp Lys Ile Lys Val
440                 445                 450                 455

CGA GAG GCC ATT GAA GCT GCA GCT ATT CCT CAC ACT TAT ATT TCT GCC     481
Arg Glu Ala Ile Glu Ala Ala Ala Ile Pro His Thr Tyr Ile Ser Ala
                460                 465                 470

AAT ATA TTT GCT GGC TAC TTG GTT GGT GGA TTA GCT CAA CTT GGT CGT     529
Asn Ile Phe Ala Gly Tyr Leu Val Gly Gly Leu Ala Gln Leu Gly Arg
            475                 480                 485

GTG ATG CCT CCT TCA GAC AAA GTA TTT CTC TAT GGA GAT GGA AAT GTC     577
Val Met Pro Pro Ser Asp Lys Val Phe Leu Tyr Gly Asp Gly Asn Val
        490                 495                 500

AAA GCT GTT TGG ATA GAT GAA GAA GAT GTT GGA ATA TAC ACA ATC AAA     625
Lys Ala Val Trp Ile Asp Glu Glu Asp Val Gly Ile Tyr Thr Ile Lys
    505                 510                 515
```

```
GCA ATT GAT GAC CCT CGC ACC CTA AAT AAG ACT GTG TAC ATC AGG CCA      673
Ala Ile Asp Asp Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro
520                 525                 530                 535

CCT TTG AAT GTT CTT TCC CAG AAG GAA GTG GTT GAA AAA TGG GAA AAA      721
Pro Leu Asn Val Leu Ser Gln Lys Glu Val Val Glu Lys Trp Glu Lys
                540                 545                 550

TTA TCA AGA AAG AGC TTG GAT AAA ATA TAT ATG TCT GTT GAG GAT TTT      769
Leu Ser Arg Lys Ser Leu Asp Lys Ile Tyr Met Ser Val Glu Asp Phe
            555                 560                 565

CTC GCA GGC ATG GAA GGT CAA TCA TAT GGA GAG AAG ATT GGA ATA TCA      817
Leu Ala Gly Met Glu Gly Gln Ser Tyr Gly Glu Lys Ile Gly Ile Ser
        570                 575                 580

CAT TTC TAT CAG ATG TTC TAT AAG GGG GAT CTT TAT AAT TTT GAA ATT      865
His Phe Tyr Gln Met Phe Tyr Lys Gly Asp Leu Tyr Asn Phe Glu Ile
    585                 590                 595

GGA CCT AAT GGA GTA GAA GCT TCC CAA CTT TAC CCA GGA GTA AAA TAC      913
Gly Pro Asn Gly Val Glu Ala Ser Gln Leu Tyr Pro Gly Val Lys Tyr
600                 605                 610                 615

ACA ACA GTG GAC TCA TAC ATG GAG CGC TAC CTA TGAAAATCTT CTTCATGAAG    966
Thr Thr Val Asp Ser Tyr Met Glu Arg Tyr Leu
                620                 625

ATATTTAAAT TCAATTTAAT GCTTTCTAAA AGTTTTTATA TTTTGACATA ATGCTAAATA   1026

TAGATGTAGA GTATCTAGAT AATAATATTC AATTGATAAT ATTCAACAAT CAGTTGAGAT   1086

GACTTTTTCC CTTTAACTGC ATGCTCAACA TATTTTATAC AAACAAGCTA ATGTCTTTTA   1146

AGGTTGAGAA ACTAAATATG GTTTTGTATT ACATGGAAAA ACCATATTTT GATATTTGAG   1206

ATTGTATTTA TTTTGAATGT TATGATTTTG ATAAAATTTG AAATTGATTA TGAACATTGT   1266

TTTAAAAAAA AAAAAAAAAA A                                              1287
```

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

```
Met Glu Glu Ser Ser Arg Ile Leu Val Val Gly Gly Thr Gly Tyr Ile
  1               5                  10                  15

Gly Arg Arg Ile Val Lys Ala Ser Ile Ala Leu Gly His Pro Thr Phe
             20                  25                  30

Ile Leu Phe Arg Lys Glu Val Val Ser Asp Val Glu Lys Val Glu Met
         35                  40                  45
```



```
Leu Leu Ser Phe Lys Lys Asn Gly Ala Lys Leu Leu Glu Ala Ser Phe
     50                  55                  60

Asp Asp His Glu Ser Leu Val Asp Ala Val Lys Gln Val Asp Val Val
 65                  70                  75                  80

Ile Ser Ala Val Ala Gly Asn His Met Arg His His Ile Leu Gln Gln
                 85                  90                  95

Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe
            100                 105                 110

Val Pro Ser Glu Phe Gly Met Asp Pro Gly Leu Met Asp His Ala Met
        115                 120                 125

Ala Pro Gly Asn Ile Val Phe Ile Asp Lys Ile Lys Val Arg Glu Ala
    130                 135                 140

Ile Glu Ala Ala Ala Ile Pro His Thr Tyr Ile Ser Ala Asn Ile Phe
145                 150                 155                 160

Ala Gly Tyr Leu Val Gly Gly Leu Ala Gln Leu Gly Arg Val Met Pro
                165                 170                 175

Pro Ser Asp Lys Val Phe Leu Tyr Gly Asp Gly Asn Val Lys Ala Val
            180                 185                 190

Trp Ile Asp Glu Glu Asp Val Gly Ile Tyr Thr Ile Lys Ala Ile Asp
        195                 200                 205

Asp Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro Pro Leu Asn
    210                 215                 220

Val Leu Ser Gln Lys Glu Val Val Glu Lys Trp Glu Lys Leu Ser Arg
225                 230                 235                 240

Lys Ser Leu Asp Lys Ile Tyr Met Ser Val Glu Asp Phe Leu Ala Gly
                245                 250                 255

Met Glu Gly Gln Ser Tyr Gly Glu Lys Ile Gly Ile Ser His Phe Tyr
            260                 265                 270

Gln Met Phe Tyr Lys Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asn
        275                 280                 285

Gly Val Glu Ala Ser Gln Leu Tyr Pro Gly Val Lys Tyr Thr Thr Val
    290                 295                 300

Asp Ser Tyr Met Glu Arg Tyr Leu
305                 310
```

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1282 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: Tsuga heterophylla cDNA PLR-Th1

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
     (A) NAME/KEY: CDS
     (B) LOCATION: 2..922

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

C AGA GTT CTA ATA GTG GGT GGC ACA GGA TAC ATA GGT AGA AAA TTT        46
  Arg Val Leu Ile Val Gly Gly Thr Gly Tyr Ile Gly Arg Lys Phe
  315                 320                 325

GTA AAA GCT AGC TTA GCT CTA GGC CAC CCA ACA TTC GTT TTG TCC AGG      94
Val Lys Ala Ser Leu Ala Leu Gly His Pro Thr Phe Val Leu Ser Arg
    330                 335                 340

CCA GAA GTA GGG TTT GAC ATT GAG AAG GTG CAC ATG TTG CTC TCC TTC     142
Pro Glu Val Gly Phe Asp Ile Glu Lys Val His Met Leu Leu Ser Phe
    345                 350                 355

AAA CAA GCG GGT GCC AGA CTT TTG GAG GGT TCA TTT GAG GAT TTC CAA     190
Lys Gln Ala Gly Ala Arg Leu Leu Glu Gly Ser Phe Glu Asp Phe Gln
360                 365                 370                 375

AGC CTT GTG GCA GCC TTG AAG CAG GTT GAT GTT GTG ATA AGT GCA GTG     238
Ser Leu Val Ala Ala Leu Lys Gln Val Asp Val Val Ile Ser Ala Val
                380                 385                 390

GCA GGA AAC CAT TTC AGA AAC CTT ATA CTT CAA CAG CTT AAA TTG GTG     286
Ala Gly Asn His Phe Arg Asn Leu Ile Leu Gln Gln Leu Lys Leu Val
            395                 400                 405

GAA GCC ATA AAA GAA GCT GGC AAC ATT AAG AGA TTT CTT CCT TCT GAA     334
Glu Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe Leu Pro Ser Glu
        410                 415                 420

TTT GGA ATG GAA CCA GAC CTC ATG GAG CAC GCT TTG GAA CCT GGT AAC     382
Phe Gly Met Glu Pro Asp Leu Met Glu His Ala Leu Glu Pro Gly Asn
    425                 430                 435

GCT GTC TTC ATT GAT AAG AGA AAG GTT CGG CGC GCC ATT GAA GCA GCA     430
Ala Val Phe Ile Asp Lys Arg Lys Val Arg Arg Ala Ile Glu Ala Ala
440                 445                 450                 455

GGC ATT CCT TAC ACG TAT GTC TCT TCA AAT ATA TTT GCT GGG TAT TTA     478
Gly Ile Pro Tyr Thr Tyr Val Ser Ser Asn Ile Phe Ala Gly Tyr Leu
                460                 465                 470

GCA GGA GGG TTG GCA CAA ATT GGC CGG CTT ATG CCT CCT CGT GAT GAA     526
Ala Gly Gly Leu Ala Gln Ile Gly Arg Leu Met Pro Pro Arg Asp Glu
            475                 480                 485

GTA GTT ATC TAT GGA GAT GGT AAC GTT AAA GCT GTT TGG GTG GAC GAA     574
Val Val Ile Tyr Gly Asp Gly Asn Val Lys Ala Val Trp Val Asp Glu
        490                 495                 500
```



```
GAT GAT GTC GGA ATA TAC ACA CTG AAA ACA ATC GAT GAT CCA CGC ACT     622
Asp Asp Val Gly Ile Tyr Thr Leu Lys Thr Ile Asp Asp Pro Arg Thr
    505                 510                 515

CTG AAC AAG ACT GTA TAT ATC AGG CCA CTC AAA AAT ATT CTC TCT CAG     670
Leu Asn Lys Thr Val Tyr Ile Arg Pro Leu Lys Asn Ile Leu Ser Gln
520                 525                 530                 535

AAG GAG CTT GTG GCA AAG TGG GAA AAA CTC TCA GGA AAG TGT TTG AAG     718
Lys Glu Leu Val Ala Lys Trp Glu Lys Leu Ser Gly Lys Cys Leu Lys
                540                 545                 550

AAA ACA TAC ATT TCT GCT GAG GAT TTT CTT GCA GGC ATC GAA GAT CAA     766
Lys Thr Tyr Ile Ser Ala Glu Asp Phe Leu Ala Gly Ile Glu Asp Gln
            555                 560                 565

CCT TAC GAA CAT CAG GTC GGA ATA TCT CAC TTC TAT CAA ATG TTT TAC     814
Pro Tyr Glu His Gln Val Gly Ile Ser His Phe Tyr Gln Met Phe Tyr
        570                 575                 580

AGT GGA GAT CTC TAT AAT TTT GAG ATT GGG CCA GAC GGT AGA GAA GCA     862
Ser Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asp Gly Arg Glu Ala
    585                 590                 595

ACA GTG CTA TAC CCT GAA GTT CAA TAC ACT ACC ATG GAT TCT TAT TTG     910
Thr Val Leu Tyr Pro Glu Val Gln Tyr Thr Thr Met Asp Ser Tyr Leu
600                 605                 610                 615

AAG CGC TAC TTA TAAGCAGGAT GAAGGTTAAT GTTCTACGAC ATGAATCCCA         962
Lys Arg Tyr Leu

CGAGAAATAC CAGAAATCTT CATTCAAGAT CAAATAATGG ATAAATAATT CAACATTAGT  1022

TCCATCAGAA ATACCAGAAA TTTCTAATCG AGTTCAAATA ATGGATAAAT AATTCATTAT  1082

TTAAGTTTTA TTTATCGAAA TAGGGCTGGA CGAATTGAAT ATATATTCAT CTGATATGGA  1142

CGGGCAGGTT GTAAAATTGC AAGCTGTACA GTAACTACGT CTTGTCGCGA AAAGCTACTA  1202

TATCGATATA ACTGATGTGA AAAGTTACCA TTTCGTAATA ACTATGCTTG AATTTATTTT  1262

TGACAAAAAA AAAAAAAAAA                                               1282

(2) INFORMATION FOR SEQ ID NO:70:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 307 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Arg Val Leu Ile Val Gly Gly Thr Gly Tyr Ile Gly Arg Lys Phe Val
1               5                   10                  15
```

```
Lys Ala Ser Leu Ala Leu Gly His Pro Thr Phe Val Leu Ser Arg Pro
            20                  25                  30

Glu Val Gly Phe Asp Ile Glu Lys Val His Met Leu Leu Ser Phe Lys
        35                  40                  45

Gln Ala Gly Ala Arg Leu Leu Glu Gly Ser Phe Glu Asp Phe Gln Ser
    50                  55                  60

Leu Val Ala Ala Leu Lys Gln Val Asp Val Val Ile Ser Ala Val Ala
65                  70                  75                  80

Gly Asn His Phe Arg Asn Leu Ile Leu Gln Gln Leu Lys Leu Val Glu
                85                  90                  95

Ala Ile Lys Glu Ala Gly Asn Ile Lys Arg Phe Leu Pro Ser Glu Phe
            100                 105                 110

Gly Met Glu Pro Asp Leu Met Glu His Ala Leu Glu Pro Gly Asn Ala
        115                 120                 125

Val Phe Ile Asp Lys Arg Lys Val Arg Arg Ala Ile Glu Ala Ala Gly
    130                 135                 140

Ile Pro Tyr Thr Tyr Val Ser Ser Asn Ile Phe Ala Gly Tyr Leu Ala
145                 150                 155                 160

Gly Gly Leu Ala Gln Ile Gly Arg Leu Met Pro Pro Arg Asp Glu Val
                165                 170                 175

Val Ile Tyr Gly Asp Gly Asn Val Lys Ala Val Trp Val Asp Glu Asp
            180                 185                 190

Asp Val Gly Ile Tyr Thr Leu Lys Thr Ile Asp Asp Pro Arg Thr Leu
        195                 200                 205

Asn Lys Thr Val Tyr Ile Arg Pro Leu Lys Asn Ile Leu Ser Gln Lys
    210                 215                 220

Glu Leu Val Ala Lys Trp Glu Lys Leu Ser Gly Lys Cys Leu Lys Lys
225                 230                 235                 240

Thr Tyr Ile Ser Ala Glu Asp Phe Leu Ala Gly Ile Glu Asp Gln Pro
                245                 250                 255

Tyr Glu His Gln Val Gly Ile Ser His Phe Tyr Gln Met Phe Tyr Ser
            260                 265                 270

Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asp Gly Arg Glu Ala Thr
        275                 280                 285

Val Leu Tyr Pro Glu Val Gln Tyr Thr Thr Met Asp Ser Tyr Leu Lys
    290                 295                 300

Arg Tyr Leu
305
```

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1328 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Tsuga heterophylla cDNA PLR-Th2

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 20..946

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

```
GAATTCGGCA CGAGCTAAC ATG AGC AGA GTT CTA ATA GTG GGT GGC ACA GGA      52
                     Met Ser Arg Val Leu Ile Val Gly Gly Thr Gly
                             310                 315

TAC ATA GGT AGA AAA TTT GTA AAA GCT AGC TTA GCT CTA GGC CAC CCA      100
Tyr Ile Gly Arg Lys Phe Val Lys Ala Ser Leu Ala Leu Gly His Pro
    320                 325                 330

ACA TTC GTT TTG TCC AGG CCA GAA GTA GGG TTT GAC ATT GAG AAG GTG      148
Thr Phe Val Leu Ser Arg Pro Glu Val Gly Phe Asp Ile Glu Lys Val
335                 340                 345                 350

CAC ATG TTG CTC TCC TTC AAA CAA GCG GGT GCC AGA CTT TTG GAG GGT      196
His Met Leu Leu Ser Phe Lys Gln Ala Gly Ala Arg Leu Leu Glu Gly
                355                 360                 365

TCA TTT GAG GAT TTC CAA AGC CTT GTG GCA GCC TTG AAG CAG GTT GAT      244
Ser Phe Glu Asp Phe Gln Ser Leu Val Ala Ala Leu Lys Gln Val Asp
            370                 375                 380

GTT GTG ATA AGT GCA GTG GCA GGA AAC CAT TTC AGA AAC CTT ATA CTT      292
Val Val Ile Ser Ala Val Ala Gly Asn His Phe Arg Asn Leu Ile Leu
        385                 390                 395

CAA CAG CTT AAA TTG GTG GAA GCC ATA AAA GAG GCT CGC AAC ATT AAG      340
Gln Gln Leu Lys Leu Val Glu Ala Ile Lys Glu Ala Arg Asn Ile Lys
    400                 405                 410

AGA TTT CTT CCT TCT GAA TTT GGA ATG GAC CCA GAC CTC ATG GAG CAC      388
Arg Phe Leu Pro Ser Glu Phe Gly Met Asp Pro Asp Leu Met Glu His
415                 420                 425                 430

GCT TTG GAA CCT GGT AAC GCT GTC TTC ATT GAT AAG AGA AAG GTT CGG      436
Ala Leu Glu Pro Gly Asn Ala Val Phe Ile Asp Lys Arg Lys Val Arg
                435                 440                 445

CGC GCC ATT GAA GCA GCA GGC ATT CCT TAC ACG TAT GTC TCT TCA AAT      484
Arg Ala Ile Glu Ala Ala Gly Ile Pro Tyr Thr Tyr Val Ser Ser Asn
            450                 455                 460
```

```
ATA TTT GCT GGG TAT TTA GCA GGA GGG TTG GCA CAA ATT GGC CGG CTT       532
Ile Phe Ala Gly Tyr Leu Ala Gly Gly Leu Ala Gln Ile Gly Arg Leu
        465                 470                 475

ATG CCT CCT CGT GAT GAA GTA GTT ATC TAT GGA GAT GGT AAC GTT AAA       580
Met Pro Pro Arg Asp Glu Val Val Ile Tyr Gly Asp Gly Asn Val Lys
    480                 485                 490

GCT GTT TGG GTG GAC GAA GAT GAT GTC GGA ATA TAC ACA CTG AAA ACA       628
Ala Val Trp Val Asp Glu Asp Asp Val Gly Ile Tyr Thr Leu Lys Thr
495                 500                 505                 510

ATC GAT GAT CCA CGC ACT CTG AAC AAG ACT GTA TAT ATC AGG CCA CTC       676
Ile Asp Asp Pro Arg Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro Leu
                515                 520                 525

AAA AAT ATA CTC TCT CAG AAG GAG CTT GTG GCA AAG TGG GAA AAA CTC       724
Lys Asn Ile Leu Ser Gln Lys Glu Leu Val Ala Lys Trp Glu Lys Leu
            530                 535                 540

TCA GGA AAG TTT TTG AAG AAA ACA TAC ATT TCT GCT GAG GAT TTT CTT       772
Ser Gly Lys Phe Leu Lys Lys Thr Tyr Ile Ser Ala Glu Asp Phe Leu
        545                 550                 555

GCA GGC ATC GAA GAT CAA CCT TAC GAA CAT CAG GTC GGA ATA TCT CAC       820
Ala Gly Ile Glu Asp Gln Pro Tyr Glu His Gln Val Gly Ile Ser His
    560                 565                 570

TTC TAT CAA ATG TTT TAC AGT GGA GAT CTC TAT AAT TTT GAG ATT GGG       868
Phe Tyr Gln Met Phe Tyr Ser Gly Asp Leu Tyr Asn Phe Glu Ile Gly
575                 580                 585                 590

CCA GAC GGT AGA GAA GCA ACA ATG CTA TAC CCT GAA GTT CAA TAC ACT       916
Pro Asp Gly Arg Glu Ala Thr Met Leu Tyr Pro Glu Val Gln Tyr Thr
                595                 600                 605

ACC ATG GAT TCT TAT TTG AAG CGC TAC TTA TAAGCAGGAT GAAGGTTAAT         966
Thr Met Asp Ser Tyr Leu Lys Arg Tyr Leu
            610                 615

GTTCTACGAC ATGAATCCCA CGAGAAATAC CAGAAATCTT CATTCAAGAT CAAATAATGG    1026

ATAAATAATT CAACATTAGT TCCATCAGAA ATATCAGAAA TTTCTAATCA AGTTCAAATA    1086

ATGGATAAAT AATTCATTAT TTAAGTTTTA TTTATTGAAA TAGGGCTGGA CGAAGCCTTT    1146

AATCAGTATT GAATATATAT TCATCTGATA TGGACGGGCA GGTTGTAAAA TTGCAAGCCG    1206

TACAGTAACT ACGTCTTGTC GCGAAAAGCT ACCATATCGA TATAACTAAG TCTTGTCGCG    1266

TAAAGCTACC ATATCGATAT AACTGATGTG ACCATTTCGT AATAACTATG CTTGTGCAGG    1326

AA                                                                   1328
```

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 309 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

```
Met Ser Arg Val Leu Ile Val Gly Gly Thr Gly Tyr Ile Gly Arg Lys
1               5                   10                  15

Phe Val Lys Ala Ser Leu Ala Leu Gly His Pro Thr Phe Val Leu Ser
            20                  25                  30

Arg Pro Glu Val Gly Phe Asp Ile Glu Lys Val His Met Leu Leu Ser
        35                  40                  45

Phe Lys Gln Ala Gly Ala Arg Leu Leu Glu Gly Ser Phe Glu Asp Phe
    50                  55                  60

Gln Ser Leu Val Ala Ala Leu Lys Gln Val Asp Val Val Ile Ser Ala
65                  70                  75                  80

Val Ala Gly Asn His Phe Arg Asn Leu Ile Leu Gln Gln Leu Lys Leu
                85                  90                  95

Val Glu Ala Ile Lys Glu Ala Arg Asn Ile Lys Arg Phe Leu Pro Ser
            100                 105                 110

Glu Phe Gly Met Asp Pro Asp Leu Met Glu His Ala Leu Glu Pro Gly
        115                 120                 125

Asn Ala Val Phe Ile Asp Lys Arg Lys Val Arg Arg Ala Ile Glu Ala
    130                 135                 140

Ala Gly Ile Pro Tyr Thr Tyr Val Ser Ser Asn Ile Phe Ala Gly Tyr
145                 150                 155                 160

Leu Ala Gly Gly Leu Ala Gln Ile Gly Arg Leu Met Pro Pro Arg Asp
                165                 170                 175

Glu Val Val Ile Tyr Gly Asp Gly Asn Val Lys Ala Val Trp Val Asp
            180                 185                 190

Glu Asp Asp Val Gly Ile Tyr Thr Leu Lys Thr Ile Asp Asp Pro Arg
        195                 200                 205

Thr Leu Asn Lys Thr Val Tyr Ile Arg Pro Leu Lys Asn Ile Leu Ser
    210                 215                 220

Gln Lys Glu Leu Val Ala Lys Trp Glu Lys Leu Ser Gly Lys Phe Leu
225                 230                 235                 240

Lys Lys Thr Tyr Ile Ser Ala Glu Asp Phe Leu Ala Gly Ile Glu Asp
                245                 250                 255

Gln Pro Tyr Glu His Gln Val Gly Ile Ser His Phe Tyr Gln Met Phe
            260                 265                 270

Tyr Ser Gly Asp Leu Tyr Asn Phe Glu Ile Gly Pro Asp Gly Arg Glu
        275                 280                 285
```

```
Ala Thr Met Leu Tyr Pro Glu Val Gln Tyr Thr Thr Met Asp Ser Tyr
    290                 295                 300

Leu Lys Arg Tyr Leu
305
```

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA probe used to isolate Forsythia intermedia dirigent protein cDNA clone

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

```
AAGGAGCTGG TGTTCTACTT CCACGACATA CTTTTCAAAG GGGATAATTA CAACAATGCC    60
ACTGCCACCA TAGTCGGGTC CCCCCAATGG GGCAACAAGA CTGCCATGGC CGTGCCATTC   120
AATTTTGGTG ACCTAATGGT GTTCGACGAT CCCATTACCT TAGACAACAA TCTGCATTCA   180
CCCCCAGTGG GTCGGGCACA AGGGATGTAC TTCTATGATC AAAAAAGTAC ATACAATGCT   240
TGGCTCGGGT TCTCATTTTT GTTCAATTCA ACTAAGTATG TTGGAACCTT GAACTTTGCT   300
GGGGCTGATC CATTGTTGAA CAAGACTAGG GACGTATCAG TCATTGGTGG AACCA        355
```

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION:"PCRprimer R20"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```
CAGCTATGAC CATGATTACG                                                20
```

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION:"PCR primer U19"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

```
GTTTTCCCAG TCACGACGT                                                 19
```

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide(NADPH) binding motif

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

```
Gly Xaa Gly Xaa Xaa Gly
1               5
```

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. An isolated protein from a lignan biosynthetic pathway selected from the group consisting of dirigent protein and pinoresinol/lariciresinol reductases.

2. An isolated protein of Claim 1 having the biological activity of dirigent protein.

3. An isolated protein of Claim 2 having the biological activity of dirigent protein from *Forsythia*.

4. An isolated protein of Claim 3 having the biological activity of dirigent protein from *Forsythia intermedia*.

5. An isolated protein of Claim 2 having the biological activity of dirigent protein from *Tsuga*.

6. An isolated protein of Claim 5 having the biological activity of dirigent protein from *Tsuga heterophylla*.

7. An isolated protein of Claim 2 having the biological activity of dirigent protein from *Thuja*.

8. An isolated protein of Claim 7 having the biological activity of dirigent protein from *Thuja plicata*.

9. An isolated protein of Claim 1 having the biological activity of dirigent protein selected from the group consisting of SEQ ID Nos: 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and 35.

10. An isolated protein of Claim 1 having the biological activity of pinoresinol/lariciresinol reductase.

11. An isolated protein of Claim 10 having the biological activity of pinoresinol/lariciresinol reductase from *Forsythia*.

12. An isolated protein of Claim 11 having the biological activity of pinoresinol/lariciresinol reductase from *Forsythia intermedia*.

13. An isolated protein of Claim 10 having the biological activity of pinoresinol/lariciresinol reductase from *Tsuga*.

14. An isolated protein of Claim 13 having the biological activity of pinoresinol/lariciresinol reductase from *Tsuga heterophylla*.

15. An isolated protein of Claim 10 having the biological activity of pinoresinol/lariciresinol reductase from *Thuja*.

16. An isolated protein of Claim 15 having the biological activity of pinoresinol/lariciresinol reductase from *Thuja plicata*.

17. An isolated protein of Claim 1 having the biological activity of pinoresinol/lariciresinol reductase selected from the group consisting of SEQ ID Nos:48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70 and 72.

18. An isolated nucleotide sequence encoding a dirigent protein.

19. An isolated nucleotide sequence encoding a dirigent protein from a *Forsythia* species.

20. A nucleotide sequence of Claim 19 encoding a dirigent protein from *Forsythia intermedia*.

21. An isolated nucleotide sequence encoding a protein having the biological activity of SEQ ID No:13 or SEQ ID No:15.

22. An isolated nucleotide sequence of Claim 19 which encodes the amino acid sequence of SEQ ID No:13 or SEQ ID No:15.

23. An isolated nucleotide sequence of Claim 19 having the sequence of SEQ ID No:12 or SEQ ID No:14.

24. An isolated nucleotide sequence encoding a dirigent protein from a *Tsuga* species.

25. A nucleotide sequence of Claim 24 encoding a dirigent protein from *Tsuga heterophylla*.

26. An isolated nucleotide sequence encoding a protein having the biological activity of SEQ ID No:17 or SEQ ID No:19.

27. An isolated nucleotide sequence of Claim 24 which encodes the amino acid sequence of SEQ ID No:17 or SEQ ID No:19.

28. An isolated nucleotide sequence of Claim 24 having the sequence of SEQ ID No:16 or SEQ ID No:18.

29. An isolated nucleotide sequence encoding a dirigent protein from a *Thuja* species.

30. A nucleotide sequence of Claim 29 encoding a dirigent protein from *Thuja plicata*.

31. An isolated nucleotide sequence encoding a protein having the biological activity of any one of SEQ ID Nos:21, 23, 25, 27, 29, 31, 33 or 35.

32. An isolated nucleotide sequence of Claim 29 which encodes the amino acid sequence of any one of SEQ ID Nos:21, 23, 25, 27, 29, 31, 33 or 35.

33. An isolated nucleotide sequence of Claim 29 having the sequence of any one of SEQ ID Nos:20, 22, 24, 26, 28, 30, 32 or 34.

34. An isolated nucleotide sequence encoding a pinoresinol/lariciresinol reductase from a *Forsythia* species.

35. A nucleotide sequence of Claim 34 encoding a pinoresinol/lariciresinol reductase from *Forsythia intermedia*.

36. An isolated nucleotide sequence encoding a protein having the biological activity of any one of SEQ ID Nos:48, 50, 52, 54, 56 or 58.

37. An isolated nucleotide sequence of Claim 34 which encodes the amino acid sequence of any one of SEQ ID Nos:48, 50, 52, 54, 56 or 58.

38. An isolated nucleotide sequence of Claim 34 having the sequence of any one of SEQ ID Nos:47, 49, 51, 53, 55 or 57.

39. An isolated nucleotide sequence encoding a pinoresinol/lariciresinol reductase from a *Thuja* species.

40. A nucleotide sequence of Claim 39 encoding a pinoresinol/-lariciresinol reductase from *Thuja plicata*.

41. An isolated nucleotide sequence encoding a protein having the biological activity of any one of SEQ ID Nos:62, 64, 66 or 68.

42. An isolated nucleotide sequence of Claim 39 which encodes the amino acid sequence of any one of SEQ ID Nos:62, 64, 66 or 68.

43. An isolated nucleotide sequence of Claim 39 having the sequence of any one of SEQ ID Nos:61, 63, 65 or 67.

44. An isolated nucleotide sequence encoding a pinoresinol/lariciresinol reductase from a *Tsuga* species.

45. A nucleotide sequence of Claim 44 encoding a pinoresinol/-lariciresinol reductase from *Tsuga heterophylla*.

46. An isolated nucleotide sequence encoding a protein having the biological activity of SEQ ID No:70 or SEQ ID No:72.

47. An isolated nucleotide sequence of Claim 44 which encodes the amino acid sequence of SEQ ID No:70 or SEQ ID No:72.

48. An isolated nucleotide sequence of Claim 44 having the sequence of SEQ ID No:69 or SEQ ID No:71.

49. A replicable expression vector comprising a nucleotide sequence encoding a protein having the biological activity of a dirigent protein selected from the group consisting of SEQ ID Nos:13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and 35.

50. A replicable expression vector comprising a nucleotide sequence encoding a protein having the biological activity of a pinoresinol/lariciresinol reductase selected from the group consisting of SEQ ID Nos:48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70 and 72.

51. A host cell comprising a vector of Claim 49.

52. A host cell comprising a vector of Claim 50.

53. A method of enhancing the expression of pinoresinol/lariciresinol reductase in a suitable host cell comprising introducing into the host cell an expression vector that comprises a nucleotide sequence encoding a protein having the biological activity of a protein selected from the group consisting of SEQ ID Nos:48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70 and 72.

54. A method of modifying the expression of pinoresinol/lariciresinol reductase in a suitable host cell comprising introducing into the host cell an expression vector that comprises a nucleotide sequence that expresses an RNA that is complementary to all or part of a nucleic acid molecule selected from the group consisting of SEQ ID Nos:47, 49, 51, 53, 55, 57, 61, 63, 65, 67, 69 and 71.

55. A method of enhancing the expression of dirigent protein in a suitable host cell comprising introducing into the host cell an expression vector that comprises a nucleotide sequence encoding a protein having the biological activity of a protein selected from the group consisting of SEQ ID Nos:13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and 35.

56. A method of modifying the expression of dirigent protein in a suitable host cell comprising introducing into the host cell an expression vector that comprises a nucleotide sequence that expresses an RNA that is complementary to all or part of a nucleic acid molecule selected from the group consisting of SEQ ID Nos:12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34.

57. A method of producing optically-pure lignans comprising introducing into a host cell an expression vector that comprises a nucleotide sequence encoding a dirigent protein capable of directing a bimolecular phenoxy coupling reaction to produce an optically pure lignan, and purifying the optically pure lignan from the host cell.

58. The method of Claim 57 wherein the nucleotide sequence is selected from the group consisting of SEQ ID Nos:12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34.

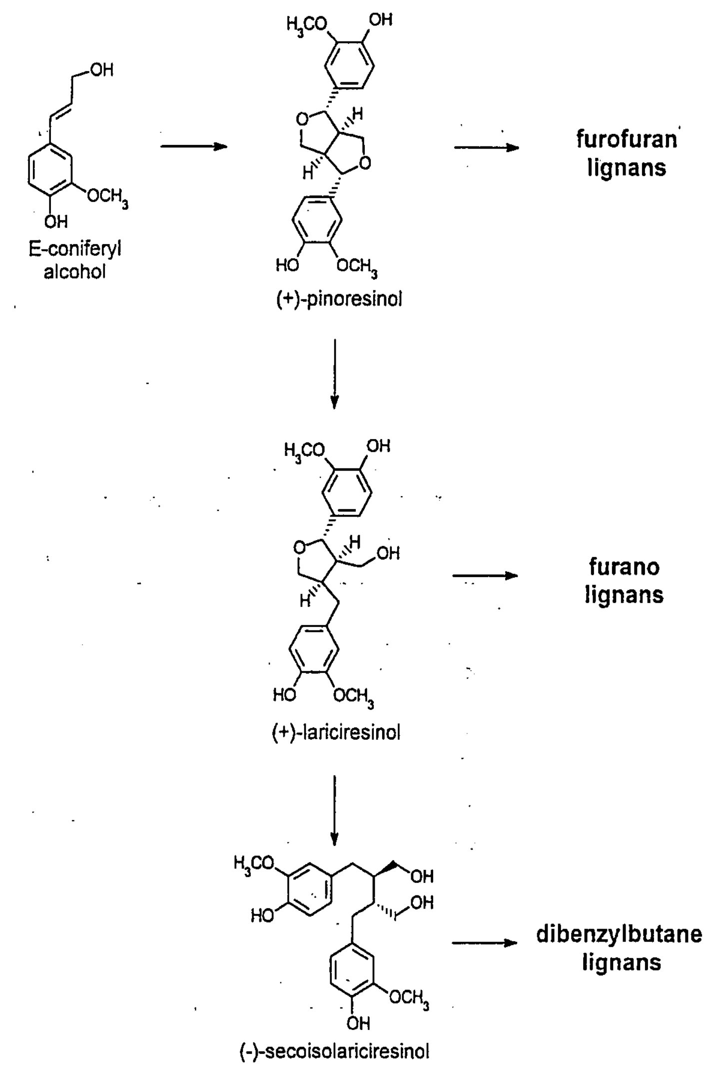

Fig. 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/20391

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C12N 9/02, 15/53, 15/29
US CL :435/189, 252.3, 325, 419, 320.1; 536/23.2, 23.6; 530/370
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/189, 252.3, 325, 419, 320.1; 536/23.2, 23.6; 530/370

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS, DIALOG
search terms: pinoresinol/lariciresinol reductase, dirigent protein, Forsythia intermedia, Thuja plicata, Tsuga heterophylla

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- A	CHU et al. Stereospecificity of (+)-pinoresinol and (+)-lariciresinol reductases from *Forsythia intermedia*. The Journal of Biological Chemistry. 25 December 1993, Vol. 268, No. 36, pages 27026-27033, see entire document.	1,10-12 ----------- 2-9, 13-58
A	KATAYAMA et al. An extraordinary accumulation of (-)-pinoresinol in cell-free extracts of *Forsythia intermedia*: Evidence for enantionspecific reduction of (+)-pinoresinol. Phytochemistry. 1992, Vol. 31, No. 11, pages 3875-3881, see entire document.	1-58

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 JANUARY 1998	23 FEB 1998

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	KEITH D. HENDRICKS
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: ____________________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.